DOCUMENT RESUME

ED 368 760 TM 021 154

AUTHOR	Vernetson, Theresa, Ed.
TITLE	Selected Papers from the Spring 1993 Breivogel Conference at the University of Florida on Alternative/Portfolio Assessment.
INSTITUTION	Florida Educational Research and Development Council, Inc., Sanibel.
PUB DATE	93
NOTE	187p.
AVAILABLE FROM	Florida Educational Research Council, Inc., P.O. Box 506, Sanibel, FL 33957 ($4 individual copies; $15 annual subscription, 10% discount on 5 or more).
PUB TYPE	Collected Works - Conference Proceedings (021) -- Tests/Evaluation Instruments (160)
JOURNAL CIT	Florida Educational Research Council Research Bulletin; v25 n1 Fall 1993
EDRS PRICE	MF01/PC08 Plus Postage.
DESCRIPTORS	Computer Assisted Testing; Curriculum Development; *Educational Assessment; Educational History; Elementary Secondary Education; *Portfolios (Background Materials); Problem Solving; Program Development; Scoring; *Standards; *Teacher Education; *Test Results; *Theory Practice Relationship
IDENTIFIERS	*Alternative Assessment; Authentic Assessment; New Standards Project (LRDC)

ABSTRACT

This edition of the "Research Bulletin" is a compilation of papers presented at the annual William F. Breivogel Conference in 1993. The conference theme was alternative and portfolio assessment. Papers were grouped into assessment in general, portfolio assessment, and alternative assessments and curriculum questions. The selected papers include: (1) "Perspectives on Alternative Assessment: What's Happening Nationally" (Thomas H. Fisher); (2) "Scoring the New Standards Project: 5 on a 6 Point Scale" (Lee Baldwin, et al.); (3) "Can Test Scores Remain Authentic when Teaching to the Test?" (M. David Miller and Anne E. Seraphine); (4) "Managing Classroom Assessments: A Computer-Based Solution for Teachers" (Madhabi Banerji and P. Charles Hutinger); (5) "Historical Roots of Current Practice in Educational Testing" (Annie W. Ward and Mildred Murray-Ward); (6) "The Portfolio: Scrapbook or Assessment Tool" (Jonnie P. Ellis, et al.); (7) "Addressing Theoretical and Practical Issues of Using Portfolio Assessment on a Large Scale in High School Settings" (Willa Wolcott); (8) "The Effective Use of Portfolio Assessment within Preservice Teacher Education: The University of Florida's Elementary Proteach Program" (Lynn Hartle and Paula DeHart); (9) "Portfolio Assessment in Teacher Education Courses" (Lyn Wagner, et al.) ; (10) "Modeling Alternative Assessment in Teacher Education Classrooms" (Mary Elizabeth D'Zamko and Lynn Raiser); (11) "An Analysis of Curriculum Domains: Implications for Assessment, Program Development, and School Improvement" (Linda S. Behar); (12) "Assessing Approaches to Classroom Assessment: Building a Knowledge/Skill Base for Preservice and Inservice Teachers" (Lehman W. Barnes and Marianne B. Barnes); and (13) "Assessing Mathematical Problem Solving from Multiple Perspectives" (Mary Grace Kantowski, et al.). (SLD)

ED 368 760

FLORIDA EDUCATIONAL
RESEARCH COUNCIL

RESEARCH BULLETIN

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)
☑ This document has been reproduced as received from the person or organization originating it.
☐ Minor changes have been made to improve reproduction quality.
● Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY
C. T. Council
TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

Selected Papers from the Spring 1993 Breivogel Conference at The University of Florida on Alternative/Portfolio Assessment

Edited by
Theresa Vernetson
Director of Extended Services
College of Education
University of Florida

TM021154

Volume 25 **Fall 1993** **Number 1**

BEST COPY AVAILABLE

FLORIDA EDUCATIONAL
RESEARCH COUNCIL, INC.

RESEARCH BULLETIN

Selected Papers from the Spring 1993 Breivogel Conference at the University of Florida
on
Alternative/Portfolio Assessment

Edited by
Theresa Vernetson
Director of Extended Services
College of Education
University of Florida

Additional copies of this book may be obtained from:
FERC, Inc.
P.O. Box 506
Sanibel, Florida 33957

Individual Copies	$4.00
Annual Subscription	$15.00

10% Discount on an order of 5 or more.
Order of less than $20.00 must be accompanied by
a check or money order.

TABLE OF CONTENTS

PART 1 9

Perspectives on Alternative Assessment: What's Happening Nationally? 11

Scoring the New Standards Project: 5 on a 6 Point Scale 17

Can Test Scores Remain Authentic When Teaching to the Test? 21

Managing Classroom Assessments: A Computer-Based Solution for Teachers 31

Historical Roots of Current Practice in Educational Testing 55

PART 2 67

The Portfolio: Scrapbook or Assessment Tool? 69

Addressing Theoretical and Practical Issues of Using Portfolio Assessment on a Large Scale in High School Settings 123

The Effective Use of Portfolio Assessement Within Preservice Teacher Education: The University of Florida's Elementary Proteach Program 133

Portfolio Assessment in Teacher Education Courses 147

PART 3 159

Modeling Alternative Assessment in Teacher Education Classrooms 153

An Analysis of Curriculum Domains: Implications for Assessment, Program Development and School Improvement 161

Assessing Approaches to Classroom Assessment: Building a Knowledge/Skill Base for Preservice and Inservice Teachers 175

Assessing Mathematical Problem Solving from Multiple Perspectives 183

Editorial Comments

This edition of the FERC *Research Bulletin* is a compilation of papers presented at the annual William F. Breivogel Conference which was held in Gainesville, on March 25, 1993. Alternative/Portfolio Assessment was the 1993 conference theme - a theme which proved to be overwhelmingly popular. Over 150 individuals participated as presenters or participants, and general agreement over the success of the conference was reached!

As can be seen from the papers presented, this conference engaged the energies and expertise of school district employees, school-based administrators and teachers, Department of Education personnel, as well as college and university professors. Not all papers presented at the Breivogel Conference are included. However, those that are included in this document cut across the important issues addressed at the conference and are divided into three parts.

Part 1 addresses the topic of assessment in general. Tom Fisher's paper describes the challenges and problems associated with alternative assessment and addresses the national assessment picture. The New Standards Project is described by Lee Baldwin, Lori Karabedian, and Candace Parks. Their description is of particular note because it includes information about the development and piloting performance-based examination system as well as the procedures and results of a local scoring conference. David Miller and Anne Seraphine explore the question of test scores as authentic assessments. Their paper includes discussion of ethical decision making, teaching test-taking skills, and asks the question, "Will authentic assessments serve the needs of accountability?" Matching assessment to district curriculum with a full-range of learner outcomes, and using technology to help teachers, by implementing an Assessment Management System (AMS), is the focus of a paper presented by Madhabi Banerji and Charles Huntinger of Pasco County Schools. Finally, drawing a close to Part 1, the history of educational testing and current practice are discussed in Annie Ward's and Mildred Murray-Ward's paper. They argue that "lessons from the past" should inform and help us develop and use truly authentic assessments.

Part 2 of the *Research Bulletin* contains the papers which deal specifically with portfolio assessment. Jonnie Ellis, Jeanette Hiebsch and Shirley Taylor present a case study of how a kindergarten-first grade Chapter 1 program changed a collection of classroom worksheets into a sound assessment tool for measuring academic growth in an instructional setting. Willa Wolcott's paper addresses theoretical issues which surfaced in a review of portfolio literature and portfolio practices. She specifically discusses the concepts of standardization versus individualization. Three papers discuss the use of portfolio assessment in preservice teacher education programs. Lynn Hartle and Paula DeHart discuss a collaborative effort between the university and cooperating teachers who supervise elementary teachers to document through portfolio assessment the students' progress toward meeting constructivist guidelines. Lyn Wagner, Ann Agnew and Dana Brock initiated portfolio assessment in two different undergraduate teacher education courses. Their paper outlines procedures for preparing portfolios which more richly and accurately portray their progress through the preservice teacher education program. The use of multiple sources of evidence is a highlight of their work. Finally, Mary Elizabeth D'Zamko and Lynne Raiser discuss a process folio approach which includes the opportunity for students to reflect on their work. They use the Foxfire strategy as a collaborative approach to teaching and learning.

Part 3 of the *Research Bulletin* deals with alternative assessments and curriculum questions. Linda Behar discusses the role of curriculum specialists and the role colleges and universities can play in assisting school districts to define professional standards and accountability measures. The elementary science curriculum and classroom assessment strategies is the topic for Lehman Barnes' and Marianne Barnes' paper. Classroom science assessment is discussed with particular attention to creating setting within which teachers can engage fully and personally in the assessment process. Assessing mathematical problem solving from multiple perspectives is the topic of a paper presented by Mary Grace Kantowski, Stephanie Robinson, Thomasenia Adams, Juli Dixon and Jeff Isaacson. They argue that "changes in assessment must include alternative techniques as well as improved paper and pencil tests: a complete assessment system, formal and informal," and they demonstrate multiple assessment measures.

The 1993 Breivogel Conference on Alternative/Portfolio Assessment was merely the beginning of conversations about what Floridians will come to know as a very challenging movement in education. This *Research Bulletin* should inform educators and the public about the first steps in tackling a monumental task. Working together with a continued focus on alternative assessments should provide rich evidence for the outcomes we wish to measure.

Theresa B. Vernetson
Director of Extended Services
College of Education
University of Florida

BOARD OF DIRECTORS

Betty Hurlbut - 1992-95
PRESIDENT
426 School Street
Sebring, FL 33870

Mary Kay Hapgood - 1993-96
PAST PRESIDENT
215 Manatee Avenue West
Bradenton, FL 34205

Lee Rowell - 1993-96
TREASURER
445 West Amelia Street
Orlando, FL 32802

Cecil Carlton - 1992-95
P.O. Box 1470
Pensacola, FL 32597

Jack McAfee - 1991-94
P.O. Box 2648
Vero Beach, FL 32960

Mahabi Banerji - 1993-96
7227 US Highway 41
Land O'Lakes, FL 33537

Tom Conner - 1991-94
P.O. Box 787
LaBelle, FL 33935

Adrian Cline - 1991-94
530 La Solona Avenue
Arcadia, FL 33821

John Headlee - 1991-94
1507 West Main Street
Inverness, FL 32650

Sandra McDonald - 1992-95
40 Orange Street
St. Augustine, FL 32084

Jayne Hartman - 1992-95
2909 Delaware Avenue
Ft. Pierce, FL 34947

Rick Nations - 1991-94
2418 Hatton Street
Sarasota, FL 33577

COUNCIL MEMBERS

County	Address	Contact Person
Alachua	620 E. University Ave. Gainesville, FL 32601	Mel Lucas
Charlotte	1445 Piatti Drive Punta Gorda, FL 33948	John Wiegman
Citrus	3741 W. Educational Path Lecanto, FL 32661	Donn Godin
Collier	3710 Estey Avenue Naples, FL 33942	Roger Otten
DeSoto	530 LaSolona Avenue Arcadia, FL 33821	Adrian Cline
Dixie	Post Office Box 4-V Cross City, FL 32628	Lloyd Jones
Escambia	Post Office Box 1470 Pensacola, FL 32597	Cecil Carlton

COUNCIL MEMBERS (Cont.)

County	Address	Contact Person
Hardee	Post Office Drawer 678 Wauchula, FL 33873	Derrell Bryan
Hendry	Post Office Box 787 LaBelle, FL 33953	Tom Conner
Hernando	919 North Broad Street Brooksville, FL 34601	Phyllis McIntyre
Highlands	426 School Street Sebring, FL 33870	Betty Hurlbut
Hillsborough	Post Office Box 3408 Tampa, FL 33601	John Hilderbrand
Indian River	Post Office Box 2648 Vero Beach, FL 32960	Jack McAfee
Lee	2055 Central Avenue Ft. Myers, FL 33901	Mike Jones
Manatee	215 Manatee Avenue W. Bradenton, FL 34205	Mary K. Habgood
Marion	Post Office Box 670 Ocala, FL 32670	Esther Oteiza
Martin	Post Office Box 1049 Stuart, FL 33494	Deana Hughes
Nassau	1201 Atlantic Avenue Fernandina Beach, FL 32034	James T. Northey
Okeechobee	100 S.W. 5th Avenue Okeechobee, FL 33472	Danny Mullins
Orange	445 West Amelia Street Orlando, FL 32802	Lee Rowell
Pasco	7227 U.S. Highway 41 Land O'Lakes, FL 33537	Madhabi Banerji
St. Johns	40 Orange Street St. Augustine, FL 32048	Sandra McDonald
St. Lucie	2909 Delaware Avenue Ft. Pierce, FL 34947	Jayne Hartman
Santa Rosa	603 Canal Street Milton, FL 32570	Richard Mancini

COUNCIL MEMBERS (Cont.)

County	Address	Contact Person
Sarasota	2418 Hatton Street Sarasota, FL 33577	Rick Nations
Suwannee	224 Parshly Street	Marvin Johns

ADVISORS

Jake Beard
College of Education
Florida State University
Tallahassee, FL 32306

Carl Balado
College of Education
University of Central Florida
Orlando, FL 32816

William Castine
College of Education
Florida A&M University
Tallahassee, FL 32307

Allen Fisher
College of Education
Florida International University
Tamiami Trail
Miami, FL 33199

John Follman
College of Education
University of South Florida
4204 Fowler Avenue
Tampa, FL 33620

Charles Hayes
Central Florida Community College
P.O. Box 1388
Ocala, FL 32678

Sam Matthews
Educational Research &
Development Center
University of West Florida
Pensacola, FL 32514

Phyllis NeSmith
Florida School Board
Association
P.O. Box 446
Nocatee, FL 33864

Rod Smith
Florida Education Center
325 W. Gaines Street, Suite 414
Tallahassee, FL 32399

Bette Soldwedel
College of Education
University of North Florida
Jacksonville, FL 32216

Theresa Vernetsen
College of Education
University of Florida
Gainesville, FL 32611

Tom Gill
College of Education
Florida Atlantic University
Boca Raton, FL 33431

Charlie T. Council
EXECUTIVE DIRECTOR
P.O. Box 506
Sanibel, FL 33957

10

Part 1

Perspectives on Alternative Assessment: What's Happening Nationally?

by

Thomas H. Fisher, Ed.D.
Florida Department of Education
Student Assessment Services Section

PERSPECTIVES ON ALTERNATIVE ASSESSMENT: WHAT'S HAPPENING NATIONALLY?

Introduction

Today's professional educational literature contains many articles describing "alternative assessment" and urging its adoption into the nation's classrooms. The underlying theme in this discussion is that educators have overused standardized achievement tests, with their reliance on multiple-choice items. Students will achieve more if they are asked to produce something instead of being required to simply respond to a limited number of choices in a multiple-choice test (Wiggins, 1989; Wolf, 1989; Stiggins, 1988; Shepard, 1989).

Teachers are urged to use other methods of educational assessment such as essays, projects, portfolios, performance tasks, and open-ended or extended exercises. Advocates of this approach anticipate a day in which instructional activities and assessment activities will be so intertwined that they cannot be distinguished. Indeed, students will actually learn from assessment exercises and will find them entertaining.

Earlier literature on this topic was written by those who enthusiastically endorse it. More recently, as researchers and psychometricians tackled the tasks of developing the new assessments, writers began to document difficulties and suggest appropriate cautions. The purpose of this paper is to describe some of the national organizations and programs which are urging the movement toward alternative assessment and to mention some of the measurement concerns identified thus far by researchers.

National Organizations Influencing the Debate

In 1989, then-President George Bush and the nation's governors met for an educational summit meeting and adopted a set of educational goals for the future. These goals were similar to those previously adopted in individual states over the last two decades but they were distinctive in that they became a tool to focus energy toward improving education in all states. The goals called for American students to be "first in science and mathematics" and to "demonstrate academic competency, responsible citizenship, and to be productive workers." President Bush carried forward with his own educational plan published as *America 2000: An American Strategy* (1991) which call for "a nine-year crusade to move us toward the six ambitious national education goals" (p.1).

A bipartisan National Education Goals Panel was created by Congress in 1990 to report on progress toward the achievement of the goals. This led to a series of reports on the issue of how one would measure progress and what criteria could be used to judge success (National Education Goals Panel, 1991a, 1991b, 1992a, 1992b).

Congress authorized the creation of the National Council on Education Standards and Testing (Public Law 102-62) and charged it with the responsibility of advising Congress on "the desirability and feasibility of national standards and tests" and to recommend "structures and mechanisms for setting voluntary education standards and planning an appropriate system of tests." The Council (1992) subsequently issued a report titled *Raising Standards for American Education* in which it concluded that standards were needed. They should include high expectations, be voluntary, and provide focus and direction. The Council further concluded that a national assessment system was technically feasible.

The U.S. Department of Labor issued its report *What Work Requires of Schools: A SCANS Report for America 2000* (1991). This document, the product of the Secretary's Commission on Achieving Necessary Skills, considered what students of today need to learn to prepare for the jobs of tomorrow. They concluded that students need to be taught to "work smarter" since unskilled jobs are disappearing. The SCANS Report encourages schools to teach students to:

identify, organize, plan, and allocate resources;
work with others;
acquire and use information;
understand complex inter-relationships; and
work with a variety of technologies.

The report does not attempt to specify a complete curriculum for education but does emphasize that students must be problem solvers and know how to learn.

As it became clear that students needed more than basic or minimal skills to succeed in future jobs, the question turned to how educational assessment and instruction could be improved. One of the earliest and most well-known efforts to define this new agenda is found in the work of the National Council of Teachers of Mathematics in its report *Curriculum and Evaluation Standards for School Mathematics*. This document sets forth a new agenda for improving mathematics instruction and defines criteria with which to judge the mathematics curriculum. The report takes the position that mathematics assessment should not exist simply to determine whether a student has a given skill but should not exist simply to determine whether a student has a given skill but should be used to determine "what students know and how they think about mathematics (p. 191)." This position strengthens the justification for using a variety of alternative assessment modes rather than just multiple-choice items.

Early exploration of the use of alternative assessment items can be attributed to the National Assessment of Educational Progress (NAEP). NAEP was authorized by Congress in 1969 and has been monitoring student achievement in a variety of subject areas since then. It currently is providing leadership in the movement toward alternative item formats by its use of performance and extended response items. Recently, NAEP expanded the proportion of such items on its tests in response to requests from the NCTM and other groups. These items have been used in Florida by virtue of the state's participation in the NAEP Trial State Assessment in grade eight mathematics. This assessment, conducted in 1990 and 1992, provided a comparison of Florida students' achievement to that of approximately 40 other states (U.S. Department of Education, 1993).

The Council of Chief State School Officers (CCSSO) in Washington, DC, recognizes that individual states face a daunting challenge in trying to move toward alternative assessment models. The Council recently authorized the State Collaborative on Assessment and Student Standards (SCASS) as a vehicle for states to cooperate and share limited resources to develop improved assessments (Roeber, 1992). Florida is participating as an observer in several projects and is a full partner in the project to develop a new approach to science assessment.

The U.S. Department of Education has made resources available for research devoted to the improvement of educational assessment methodology. Among other things, it funded the National Center for Research on Evaluation, Standards, and Student Testing at the University of California. This project is devoted to conducting research into various types of alternative (or "authentic") assessment methods. Several articles and research reports have been made available on this topic (National Center for Research on Evaluation, Standards, and Student Testing, 1992).

Private foundations also have provided research monies to develop new assessment methods. One of the prominent studies is the New Standards Project coordinated by the Learning Research and Development Center at the University of Pittsburgh and the National Center on Education and the Economy (1992). This project hopes to create prototype educational standards and assessment exercises which will be instructionally sound, can be scored by local teachers, and will be entertaining for the students. Florida has three school districts which are participating in the New Standards Project: Orange, Indian River, and Charlotte counties.

Problems of Terminology

There appears to be inconsistency of terminology in the discussion about alternative assessment. Some authors refer to authentic assessment (Wiggins, 1989) or to performance assessment (Stiggins, 1987). Regardless of the word chosen, there appears to be a common theme—namely, that students should be assessed by observing things they can do.

Mitchell (1992) offers the following definition for assessment:

> "Assessment is an activity that can take many forms, can extend over time, and aims to capture the quality of a student's work or of an educational program."

Given this definition, it is clear that an "assessment" is not necessarily identical with a simple "test." It implies that an assessment includes things which students do to demonstrate that they have achieved a given outcome. It is this variety that provides the opportunity for a rich and varied assessment and instruction interaction, and it is also the feature which makes the psychometric task so challenging. With only a moment's reflection, one can realize it is more difficult to score a research project than to score a 50-item multiple-choice test.

Another terminology problem which has emerged is in the use of the word "standard." Educational Testing Service (1992) published a report titled *National Standards for Education: What They Might Look Like*. The report mentions four different definitions for this term:

1. A clear statement of what students should know . . . at points in their schooling.
2. Performance levels that students should be able to attain.

3. Specifications and definition of the necessary and desirable core of knowledge in a subject area.
4. Achievement of a particular point on a performance scale or passing score on a test.

Thus, to avoid confusion when engaging in discussions about improvements in educational assessment, educators should be certain they have shared meanings for important terms.

Technical Considerations

When a classroom teacher provides instruction and evaluates student progress, there is a general assumption that the teacher is competent to complete both tasks. The teacher's judgment is not ordinarily questioned, except where the student may feel that the teacher was "unfair."

However, when one operates in a larger context as in program evaluation or district and state educational assessments, more attention is paid to whether the assessment exercises and scoring mechanisms are properly designed and implemented. This is particularly true whenever there are high stakes involved such as holding the student or school accountable for meeting educational standards.

In the initial stages of discussion about alternative assessment, the focus was on the need for improved assessment practices and on the faults of continued use of multiple-choice achievement tests. Little attention was paid to the measurement pitfalls surrounding alternative assessment procedures. In recent months, articles have appeared which document some of the difficulties with alternative assessment strategies and propose solutions to these problems. See, for example, Williams, Phillips, and Yen, 1991; Beck, 1991; Miller and Legg, 1991; Koretz, et al, 1992.

Generally, any assessment exercise must generate scores which can be validly interpreted, be reliable, and free from bias against any specific group of students. Alternative assessment exercises tend to be interesting to students and teachers; however, this is not synonymous with psychometric respectability.

Several issues are emerging as central in the discussion about the practicality of alternative assessment exercises. First, there is the question of whether the results from a few alternative exercises, no matter how "rich" they are, can be generalized to the domain of interest. With traditional achievement tests, from 40 to 100 items might be used to sample broadly across the curriculum of interest (e.g., ninth grade mathematics). With alternative approaches, because so much time is required of administration, the student may only take a few exercises, making it difficult if not impossible to generalize the student's proficiency.

A second related issue is determining how many scorers will be needed to reliably judge alternative assessment exercises. Clearly, if only one rater is used, there is no cross-verification that the rater correctly applied the scoring criteria. If more than one rater is used, reliability will increase, but so will scoring costs. Compromises between practicality and measurement precision will be necessary.

The impact of this problem has been documented by Dunbar, Loretz, and Hoover (1991). The authors provide charts demonstrating the impact on score reliability when multiple readers are used for a given task. The extant research studies show curvilinear trends which stabilize after an exercise has been read by three or four readers. Similarly, the authors demonstrate that reliability is a function of the number of tasks given students, with stability not achieved until ten or more tasks have been presented.

Third, one has to consider whether it is possible to write equivalent alternative exercises year after year so successive groups of students can be measured by the same yardstick. There is no question that an infinite number of interesting mathematics exercises can be created—the problem is how one would determine that the exercises elicit the same type of behaviors from the students. Without a method for making this determination, one might inadvertently administer harder or easier exercises to students thus confounding longitudinal measurements.

The last major difficulty appears to be in the amount of resources needed to implement alternative assessment exercises. The cost of a 150 item multiple-choice test which is scanned by optical readers might be less than $2.00 per student. The cost of a single 45-minute essay holistically scored is approximately $4.00 to $5.00 per student. If alternative exercises are implemented in several subject areas, the costs can easily run anywhere from $10.00 to $100.00 per student. In addition, since alternative assessment exercises are supposed to be interwoven with the actual instruction, considerably more time is needed for administration. Typically, the teacher might spend five consecutive days preparing for a lesson, presenting the lesson, administering the exercises, and debriefing students afterwards.

Conclusions

Alternative assessment procedures have much to offer education. They provide the opportunity for teachers to improve classroom instruction and evaluation; they provide interesting and challenging exercises for students to complete. This movement can become a cornerstone of our efforts to make sweeping improvements in education.

In moving forward in this arena, educators would be well-advised to proceed with deliberation. Understand your purposes and pay attention to technical issues. Build models of your assessment strategies and pilot-test them. Ask whether you can defend the new assessment strategy if challenged by a student who believes the scoring procedure are unfair.

Involve the very best people you can in the developmental processes, and anticipate making changes and improvements as you move ahead. Do not consider alternative assessment approaches to be a simple solution to education's present difficulties. Your best bet is to create a mixture of assessment strategies, some new and some traditional, so that instruction can be improved while maintaining valid and reliable measures for group and program evaluation purposes.

REFERENCES

Beck, M. (1991). *Authentic Assessment for Large-Scale Accountability Purposes: Balancing the Rhetoric.* Paper presented at the annual meeting of the American Education Research Association, Chicago, IL.

Educational Testing Service (1992. *National Standards for Education: What They Might Look Like.* Princeton, NJ: Author.

Dunbar, S.; Koretz, D.; & Hoover, H. (1991). Quality Control in the Development and Use of Performance Assessments. *Applied Measurement in Education,* 4 (4), 289-303.

Koretz, D.; McCaffrey, D.; Klein, S.; Bell, R.; & Stecher, B. (1992). *The Reliability of Scores from the 1992 Vermont Portfolio Assessment Program.* Los Angeles, CA: RAND Institute on Education and Training and the National Center for Research on Evaluation, Standards, and Student Testing at the University of California.

Learning Research and Development Center at the University of Pittsburgh and the National Center on Education and the Economy (1992). *The New Standards Project, 1992-95: A Proposal.* Pittsburgh, PA and Rochester, NY: Authors.

Miller, D. & Legg, S. (1991). *Alternative Assessment in a High-Stakes Environment.* Tallahassee, FL: Florida Department of Education.

Mitchell, R. Testing for Learning: from Testing to Assessment. *Perspective,* 4, (Spring 1992).

National Center for Research on Evaluation, Standards, and Student Testing (1992). *The CRESST Line.* Los Angeles, CA: Author.

National Council on Education Standards and Testing (1992). *Raising Standards for American Education.* Washington, DC: Author.

National Council of Teachers of Mathematics (March 1989). *Curriculum and Evaluation Standards for School Mathematics.* Reston, VA: Author.

National Education Goals Panel (1992a). *Executive Summary: The National Education Goals Report.* Washington, DC: Author.

National Education Goals Panel (1992b). *Executive Summary: The National Education Goals Report.* Washington, DC: Author.

National Education Goals Panel (1991a). *Measuring Progress Toward the National Education Goals: A Guide to Selecting Indicators.* Washington, DC: Author.

National Education Goals Panel (1991b). *Measuring Progress Toward the National Education Goals: Potential Indicators and Measurement Strategies.* Washington, DC: Author.

Shepard, L. (1989, April). Why We Need Better Assessments. *Educational Leadership,* 46 (7), 1-17.

Stiggins, R. (1987). Design and Development of Performance Assessments. *Educational Measurement: Issues and Practice,* 6(3), 33-42.

U.S. Department of Education (1991). *America 2000: An Education Strategy.* Washington, DC: Author.

U.S. Department of Education (1993). *NAEP 1992: Mathematics State Report for Florida.* Washington, DC: Author

U.S. Department of Labor. *What Work Requires of Schools: A SCANS Report for America 2000* (1991). Washington, DC: Author.

Wiggins, G. (1989). A True Test: Toward More Authentic and Equitable Assessment. *Phi Delta Kappan,* 70 (9), 703-713.

Williams, P.; Phillips, G.; & Yen, W. (1991). *Measurement Issues in High Stakes Performance Assessment.* Paper presented at the annual meeting of the American Educational Research Association, Chicago, IL.

Scoring the New Standards Project: 5 on a 6 Point Scale

by

Lee Baldwin, Lori Karabedian, Candace Parks
Orange County Public Schools

SCORING THE NEW STANDARDS PROJECT: 5 ON A 6 POINT SCALE

Background

The New Standards Project is a joint effort of the Learning Research and Development Center of the University of Pittsburgh and the National Center on Education and the Economy. The goal of the New Standards Project is to develop a national performance-based examination system to use as a means to evaluate students, teachers, schools and systems against national standards in several subject areas. The examination components include performance assessments, portfolios and projects. Performance assessments, which take place over a period of hours or days and focus on thinking, problem solving and the capacity to apply knowledge to complex real-life problems, have been designed and piloted.

The performance assessments were developed in the areas of reading/writing and mathematics at the fourth grade level in the 1991-92 school year. After the tasks were refined, they were piloted in April in fourth grade classrooms in Indian River and Orange counties. Five teachers in reading/writing and five teachers in mathematics in each county conducted the pilots.

A national scoring conference was held during the summer to design and refine rubrics to be used in holistic scoring of the performance assessments. A focus of the national conference was to determine the feasibility and accuracy of scoring a large scale assessment in order to validate this approach to assessment.

States were then encouraged to conduct local scoring conferences to validate the consistency of the scoring procedures used at the national scoring conference. In August 1992, a local scoring conference was held in Orlando, Florida. The purpose of this conference was to establish reliable , holistic scoring procedures at the state level. The holistic scoring procedure produces one score for a student's work on a task. Assessing holistically focuses the scores on the total product, including the thinking process demonstrated by the student as well as the final response.

Purpose

The purpose of this paper is to report the procedures and results of the local conference. Specific questions of interest include the following:

1. To what extent do the scores of raters agree?
2. How do local results compare to results from the national scoring conference.
3. Is there a relationship between task difficulty and rater agreement?

Method

Participants included twelve teachers and two district-level testing administrators from Indian River County, Palm Beach County and Orange County, Florida. The group worked with three reading tasks, three writing tasks and four math tasks. The group was split into two scoring teams, one containing six members and one containing eight members. An equal number of scorers was required for each tram so that each member would have a partner for double scoring of papers.

Each team had a leader who oriented the raters to the rationale behind holistic scoring and the scoring procedures. All raters then read the text and directions for the first task. The team leader then reviewed and discussed the rubric with the group. The rubric for reading was based on a four-point scale, and the rubrics for math and writing were based on a six-point scale. Each rater received a set of benchmark papers with each point on the rubric represented. The group scored the benchmark papers for each level of the rubric and discussed results.

The first set of real papers was then scored. After three papers were scored, all scorers traded with their partner and scored their partner's papers to check for the degree of consistency between the two raters. If there were any concerns with the results, a group discussion was held. Negotiation between raters occurred if their scores did not agree. If agreement could not be reached after negotiation, a third rater was asked to score the paper to mediate a final score. Raters then continued scoring the remainder of the real papers, recording scores and identification numbers on the score sheet. After scoring each set of papers, the rater traded with his or her partner for double scoring. The group would then debrief about the rubric and student responses and make suggestions about refining the tasks and suggestions for future scoring conferences. This process was completed with each new task.

Data Analysis

SAS 6.07 was used to analyze the data by task. The following statistics were computed for each task and are reported in Table 1:

1. Percent perfect agreement - the percent of cases in which the score of Rater 1 matched the score of Rater 2 without negotiation.

2. Percent of agreement +/-1. The percent of cases in which the score of Rater 1 was within one point of the score of Rater 2 without negotiation.

3. Correlation before - correlation between scores of Rater 1 and Rater 2 before negotiation.

4. Correlation after - correlation between scores of Rater 1 and Rater 2 after negotiation.

5. Rater 1 $\bar{x}$ - mean score for Rater 1
   Rater 2 $\bar{x}$ - mean score for Rater 2
   Overall $\bar{x}$ - mean score for the task

The percentage of agreements and correlations were computed to determine interrater reliability. Mean scores were computed to assess relative degree of difficulty for each task.

Table 1 Interrater Reliability and Difficulty Indicators By Task

Task	N	% Perfect Agreement	% Agreement ±1	Corr. Before	Corr. After	Rater 1 $\bar{x}$	Rater 2 $\bar{x}$	Overall $\bar{x}$
READING								
Wolves	52	56	98	.71	.78	2.08	2.18	2.13
Camels	93	50	97	.60	.78	2.15	2.17	2.16
Folktales	67	67	100	.86	.95	2.33	2.34	2.34
	AVG.	58	99					
WRITING								
Memories	60	50	90	.74	.85	2.53	2.52	2.53
Wolves	52	60	97	.83	.94	2.77	2.75	2.76
Camels	25	52	98	.74	.82	2.80	2.84	2.82
	AVG.	54	92					
MATH								
LEP Buns	52	51	97	.84	1.0	2.95	2.95	2.95
Pizza Party	60	43	90	.79	.98	3.0	2.98	2.98
Buns	19	100	100	1.0	-	3.32	3.47	3.39
Building with Tiles	60	42	94	.76	1.0	4.08	4.08	4.08
	AVG.	59	95					

Discussion

Question 1: To what extent do raters agree?

In reading, raters were in perfect agreement or within one point of each other 99 percent of the time. Raters assigned identical scores in an average of 58 percent of the cases. In another 41 percent of the cases, scores of the two raters differed by only one point. Consequently, correlation values were very high, ranging from r=.60 to r=.86 before negotiations and from r=.78 to r=.95 after negotiations.

In writing, raters were in perfect agreement or within one point of each other 92 percent of the time. Raters assigned identical scores in an average of 54 percent of the cases. In another 38 percent of the cases, scores of the two raters differed by only one point. Again, correlation values were very high, ranging from r=.74 to r=.83 before negotiations and from r=.82 to r=.94 after negotiations.

In math, raters were in perfect agreement or within one point of each other 95 percent of the time. Raters assigned identical scores in an average of 59 percent of the cases. In another 36 percent of the cases, scores of the two raters differed by only one point. Correlation values were highest in math, ranging from r=.76 to r=1.0 before negotiations and from r=.98 to r=1.0 after negotiations.

Question 2: How do results compare to results from the National Scoring Conference?

Table 2 reports correlation coefficient ranges from both the national and local scoring conferences (national figures are from: Preliminary Results from the New Standards Project Big Sky Scoring Conference, Lizanne DeStefano, University of Illinois at Urbana, Champagne). It can be seen that the values for the local conference are significantly higher. This is not unexpected as many modifications and enhancements of the rubrics took place at the national conference. In addition, local raters benefited from the experience of those who participated in the national conference and were able to further refine the scoring procedures.

Table 2 Comparison of Correlation Coefficient Ranges from National and Local Scoring Conferences

	National	Local
Reading	.22 - .74	.78 - .95
Writing	.58 - .62	.82 - .94
Math	.25 - .65	.98 - 1.0

Question 3: Is there a relationship between task difficulty and rater agreement?

In reading, it appears that the easier the task (as indicated by a higher mean score), the higher the agreement between raters. However this does not hold true for writing and math. Perhaps the relationship in reading is a function of the fact that it is measured on a four-point rubric as opposed to the six-point rubric used in writing and math. In addition, the reading tasks appear to be more equal in difficulty than either the writing or math tasks. The difference between the highest and lowest means is .21 in reading, .29 in writing and 1.13 in math.

Conclusion

The local scoring conference served to corroborate the findings of the national conference and further demonstrate that rater reliability could be improved based on refinements that occurred at the national and local level. It was further shown that a group with little or no knowledge of performance assessment and holistic scoring could be brought together and trained to be accurate and consistent raters. We would conclude that given clear tasks, rigorous training and well-defined scoring procedures, holistic scoring can be accurate and reliable. Finally, as tasks and scoring rubrics are further refined, the scoring procedures will continue to improve.

CanTest Scores Remain Authentic when Teaching to the Test?

by
M. David Miller
and
Anne E. Seraphine
University of Florida

Abstract

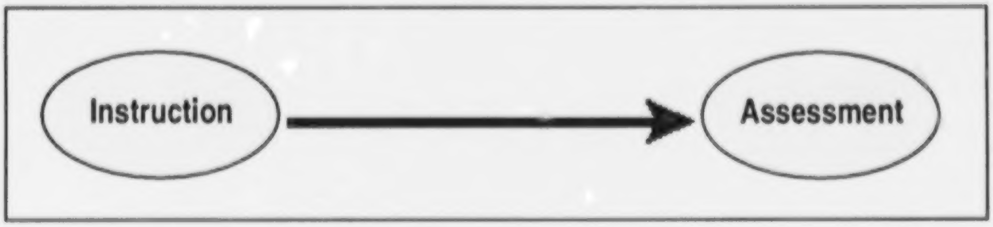
Although some educators have suggested authentic tests as a solution to the problem of artificially inflated scores from teaching to paper and pencil tests, we argue that teaching to test under high-stakes conditions could be more problematic with the new forms of assessment. The wide range of methods that can be potentially used in authentic assessments introduce a method variance that is not part of the construct to be measured. As a consequence, teaching the specific methods used in the assessment potentially invalidates the uses and interpretations that can be made from the test scores by narrowing the definition of the construct measured.

Can Test Scores Remain Authentic when Teaching to the Test?

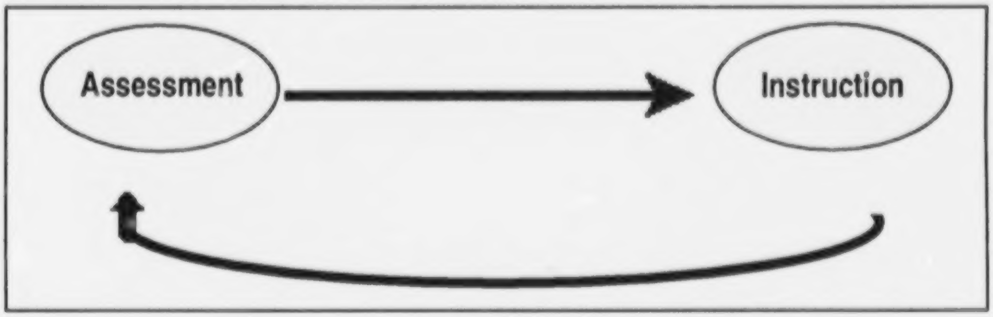
The link between instruction and assessment is at the heart of authentic assessment. Authentic assessments are intended to more closely mirror the teaching and learning process, resulting in greater instructional fidelity for the tests. Thus, the role of tests, particularly in low-stakes assessments, can be seen as in Figure 1a, with instruction preceding and determining the methods, skills, and content of the assessment.

Figure 1.

a. Traditional link of assessment and instruction (low stakes).



b. MDI link of assessment and instruction (high stakes).



In high-stakes assessments, the link of instruction and testing is viewed in a different way. Proponents of traditional measurement driven instruction (i.e., paper and pencil assessments) as well as proponents of authentic assessments have suggested assessment can be implemented not to mirror instruction but to guide instruction and learning. For example, Wiggins (1989) argued that reform in education - methods of teaching and the goals of student learning - can best be accomplished through a change in assessment, because tests "determine what teachers actually teach and what students actually learn" (p. 41). More authentic tests that replicate the core of the academic disciplines, through a reliance on methods and processes used to effectively learn each discipline, are needed for the reform movement to realize its goals. These new forms of assessment can then be used to clarify and set standards of academic excellence. Given these conditions, Wiggins suggested that testing will again serve teaching and learning. In fact, with these changes in assessment, it is no longer inappropriate to teach to the test; instead, "we should 'teach to the test' " (p. 41).

However, this view of assessment does not follow the model shown in Figure 1a. Instead, the assessments will precede instruction and learning and, as a result, determine the methods, skills, and content that are taught and learned. Thus, the link of assessment and instruction needs to be conceptualized as bidirectional as in Figure 1b rather than unidirectional as in Figure 1a. This feedback loop of testing and instruction is known to narrow instructional practices as well as the construct interpretation (i.e., validity) of the assessment in traditional assessments.

In this paper, we argue that the feedback loop will have the same effect of narrowing the construct interpretation with authentic assessment. In fact, the inherent variance across methods of operationalizing authentic assessments, which is not present with traditional assessments, may make construct interpretation

BEST COPY AVAILABLE

more problematic in the high-stakes situation. We argue that the context bound nature of authentic assessments will result in specific instructional practices that are likely to threaten the validity as well as the generalizability of the assessments.

In the first section, we show that the literature suggesting the utility of authentic assessment in a high-stakes environment is parallel to the literature on measurement-driven instruction. The aim of each is to use assessment to shape teaching practices as in Figure 1b. While traditional assessments tend to lead to low-level learning, authentic assessment will result in contextualized learning that may not lead to a generalizable construct.

In the second section, we discuss the range of teaching practices that can be engaged in to prepare for a specific test. Furthermore, we argue that teaching test-taking skills, which is not a problem for traditional forms of assessment, is a problem for authentic assessments, because it threatens both construct interpretation and generalizability.

And finally in the third section, we examine the threats to the meaningfulness and generalizability of test score interpretations more closely. First, we examine the effect of the feedback loop on the construct interpretation when instruction has been narrowed as a result of the assessment. Second, the generalizability of the assessments, which are lauded for their context bound nature, is examined.

High Stakes and Authentic Assessment

According to Wiggins, properly designed authentic forms of assessment allow " 'accountability' to serve student learning" (p. 45); a similar claim was made by advocates of more traditional forms of assessment. Popham (1987) suggested that measurement driven instruction (MDI) would result in educational reform only if it were "properly conceived and implemented" (p. 680). His suggestions were as follows: (a) criterion-referenced tests should be used, (b) the tests should include only defensible content, (c) only a manageable number of targets should be tested, (d) the tests should be designed to illuminate instruction, and (e) instructional support should be provided for teachers to teach the targeted content.

Recently, Popham (1992) amended his earlier position to advocate a shift in the type of assessment that could be effectively used for MDI. These new forms of assessment differ from the tests proposed for MDI in terms of the degree of specificity for the test specifications with the new tests falling somewhere between rigidly specified criterion-referenced tests and more globally specified norm-referenced tests. The mid-level detail in test specifications is characterized by a brief verbal description of the skills and illustrative, but nonexhaustive, items. These specifications should be "amendable to the delineation of multiple, not single, assessment tactics" (p. 17).

In line with Popham's most recent guidelines for improved MDI are the suggested guidelines for properly designed authentic assessments. Wiggins (1989) recommended that these forms of assessments should share four fundamental characteristics:

> First, they are designed to be truly representative of performance in the field... Second, far greater attention is paid to the teaching and learning of criteria to be used in the assessment. Third, self-assessment plays a much greater role than in conventional testing. And fourth, the students are often expected to present their work and defend themselves publicly and orally to ensure that their apparent mastery is genuine (Wiggins, 1989, p. 45).

Clearly, these criteria are simply extensions of Popham's most recent recommendations; contrasts between the two sets of guidelines can be seen primarily in the recommended level of student involvement in the assessment process. As compared to more traditional forms of MDI, authentic assessments purportedly allow a greater degree of self-assessment on the part of the student. Otherwise, it appears authentic assessment is just one more form of MDI, albeit more context bound.

Traditional forms of MDI have led to a narrowing of the curriculum, that is an overuse of instructional practices that engender only lower level cognitive processing. While the initial intent of high-stakes authentic assessments are to expand the curriculum, the assessments will be limited by time constraints and the particular methods of assessment chosen to represent the larger domain. As a result of the limited range of methods used in the assessment and the structuring of the curriculum to match the assessment (Wiggins' second characteristic), the curriculum may fail to draw from the larger domain. Thus, the range of teaching practices to match the curriculum to the assessment need to be considered to understand the effect of authentic assessments on instruction.

Test-Related Teaching Practices

Teaching to the test encompasses a wide range of activities that the teacher can use to improve test scores for authentic assessments as well as for more traditional forms of assessment. Mehrens and Kaminski (1989) provided a framework comprised of seven levels of test preparation activities, or teaching practices specifically tied to the test:

1. general instruction on objectives not determined by looking at the objectives measured on standardized tests;
2. teaching test-taking skills;
3. instruction on objectives generated by a commercial organization where the objectives may have been determined by looking at objectives measured by a variety of standardized tests . . .;
4. instruction based on objectives (. . .) that specifically match those on the standardized test to be administered;
5. instruction on specifically matched objectives (. . .) where the practice (instruction) follows the same format as the test questions;
6. practice (instruction) on a published parallel form of the same test; and
7. practice (instruction) on the same test (p. 16).

The seven levels of test preparation can be placed on a continuum of ethicality. In this case, level one activities would be considered ethical, while level seven activities would be considered unethical; the ethicality of activities falling between the two endpoints would be somewhat debatable, being dependent on the testing situation.

According to Mehrens and Kaminski (1989), the shift from ethical to unethical practices occurs between (3) and (5). We propose, however, that the point at which this shift occurs should be reconsidered in the case of authentic assessments.

Rather than thinking of this continuum as measuring the ethics of test preparation practices, or teaching to the test, the larger issue may be how each of these practices effects the generalizability of the construct being measured. If the activity increases test scores without a commensurate change in the measured construct, or if the consequences of the activity is to create a shift in the definition of the construct (typically narrowing the interpretation), the teaching practice should not be used. On the other hand, if the validity of the construct is maintained, the teaching practice would be appropriate.

Given this conception of appropriateness of practices for teaching to the test, clearly level (3) and above are inappropriate for authentic assessments which are intended to measure higher-order thinking in **novel** situations. However, even level (2) — teaching test-taking skills — may be inappropriate in authentic assessments, since test-taking skills are necessarily linked to some method of assessment which is part of the novelty of the assessment. In traditional assessments, the limited number of methods used gives all students the same advantage in obtaining training on test-taking skills. However, with authentic assessments, the match of instruction and the method of assessment plays a larger role because of the range of methods available in the assessment and in the instruction. Students who receive more instruction on the particular method would score higher without a commensurate gain in the construct.

Construct Validity

Linn, Baker, and Dunbar (1991) suggested eight criteria that should be used in any "serious validation" of alternative assessments: intended and unintended consequences of test use, fairness, transfer and generalizability, cognitive complexity, content quality, content coverage, meaningfulness, and cost and efficiency. Traditional validation techniques may be inadequate for the validation of authentic tests, and authentic assessments may not measure up on these criteria.

Although all of the above dimensions should be examined in a validation study, the consequences of test use, including teaching to the test, should receive a special emphasis when authentic, as well as more conventional forms of assessment, are used for accountability purposes (Miller & Legg, in press). High-stakes testing places pressure on students, teachers, and administrators to do well, resulting in practices which may alter the definition and generalizability of the measured construct. The end result is the narrowing of the construct, which violates at least five of the eight criteria recommended at the beginning of this section (i.e., transfer and generalizability, cognitive complexity, content quality, content coverage, and meaningfulness). Thus, the consequences of teaching to the test in a high-stakes testing program has a large effect on the validity of the test scores (Messick, 1989). In particular, we examine the effect of teaching to the test on the interpretation of the construct (cognitive complexity, content quality and content coverage), the relation of the construct to the particular methods of measurement (meaningfulness,) and the generalizability of the assessment (transfer and generalizability).

Construct interpretation

Airasian (1988) suggested that it is unlikely that high-stakes testing, alone, will lead to educational reform, because stakes (high or low) is only one of the factors that influences the extent to which measurement affects instruction. Instead, "stakes, standards, and content work together to influence or not influence the instruction presented to pupils" (Airasian, 1988, p. 7). Airasian (1988) claimed that, in addition to a high-stakes environment, high standards and curriculum specific content are required to ensure MDI. However, he warns us that even under optimal conditions, measurement driven instruction can go awry, because of the absence of clearly delineated instructional strategies that foster the development of higher order thinking in students. In fact, the relevance of these warnings is heightened within the context of authentic testing.

Proponents of authentic assessment have claimed that one of the chief merits of this form of assessment is its ability to measure higher order thinking skills and as a result promote instructional strategies aimed toward the development of complex cognitive processing such as the ability to formulate hypotheses, seek alternatives and make judgments, and to monitor and control one's problem-solving strategies. However, "no clear body of knowledge exists regarding the conceptualization of higher level behaviors such as reasoning, critical thinking, and application, nor are there well-validated instructional strategies to teach such higher level processes" (Airasian, 1988, p. 9).

Furthermore, a lack of well-validated instructional techniques may result in the corruption of the very construct measured by MDI tests that emphasize higher level skills (Airasian, 1988) in two ways. First, if there are no valid instructional techniques available to teach higher order thinking skills, then skills learned outside of the formal instructional setting will be tested. As a result, these tests will be measures of "general abilities" (p. 10) rather than of the achievement construct. Second, in the absence of appropriate strategies that engender complex cognitive processing, teachers in a high-stakes environment will feel pressured to teach "to known test items in a rote manner with little hope of generalizability to other item types" (Airasian, 1988, p. 10).

In sum, one possible consequence of using authentic tests to foster the instruction of higher level thinking skills, or measurement driven instruction, is that the test may no longer measure the construct that it was originally intended to measure. Teaching to the test, even within the context of authentic assessment, may pose a serious threat to the validity of these forms of assessment; that is, the uses and interpretations made of the test scores will be distorted. As Linn (1983) cautioned, inferences made from test scores become "suspect when the match [between instruction and the test] becomes too close" (p. 187).

The confounding of authentic constructs and methods

The higher level learning processes associated with any particular academic discipline are complex and diverse. As a result, a wide range of methods can be used to teach and learn an academic discipline, while the assessment must select a single, or a limited number, of methods to measure the construct. Given that the goal of the assessment is to measure competence in the academic discipline, emphasis on any single method of assessment in teaching test-taking skills could increase scores only when the particular method is used and, consequently, narrow the interpretation that can be attached to a score.

For example, Baxter, Shavelson, Goldman and Pine (1992) used both ratings of notebooks and direct observations to measure hands-on science achievement. A teacher could teach students methods for recording observations and organizing entries in a notebook. Although scores would increase on the notebooks, hands-on science achievement would not increase as measured by direct observation or other forms of assessment. As a result, the generalizability of the interpretation from the notebooks is narrowed from "hands-on science" to "hands-on science as measured by notebooks." Thus, the range of questionable or inappropriate test preparation activities expands simultaneously with the growth in methods, or forms, of assessment. Practices that are acceptable with a single method, or a limited number of methods, of assessment, such as teaching test-taking skills when only one form of assessment is used (i.e., multiple-choice tests), become inappropriate with a wider range of assessment methods because the practices are too closely linked with the particular test restricting the interpretation of the test score. Thus, the validity of test scores is adversely effected when teaching to the test is limited to the measured sample of tasks rather than the domain defined by the construct (Mehrens, 1992).

On the other hand, this does not imply that students should not be taught how to record observations and organize entries in a notebook. Since the skills are a part of the construct being measured, teachers should teach and test these types of skills. However, the appropriateness of this teaching and testing within a limited range of methods is linked to the intent of the teaching activity and the testing. When the teacher's intent is to teach notebook skills as a part of the larger curriculum, or as a part of a wider range of methods used in assessment, the teaching practice is appropriate and fits into level (1) above - general instruction on objectives not determined by the standardized test. In addition, classroom assessment using the same form of testing is entirely appropriate for measuring teaching and learning because the construct being measured is more narrowly defined and interpreted. That is, did the students learn how to use notebooks effectively in recording hands-on science tasks? However, when assessment is for accountability purposes, the construct (e.g., hands-on science) is more broadly defined and not necessarily linked to any particular form of assessment. Under these conditions, teaching specific test-taking skills (e.g., notebook skills), simply to increase scores on the particular form of the accountability assessment, is inappropriate.

It follows that the appropriateness of teaching activities, which increases test scores, is intricately linked to (a) the level of the teaching practice on the continuum above, (b) the intent of the teaching practice (increasing test scores *vs.* learning), and (c) the intent of the assessment (accountability *vs.* classroom testing). As discussed above, the same test preparation activity shifts from being inappropriate when using an accountability test, to being appropriate when using a classroom assessment. The common element in deciding upon the appropriateness of a teaching activity linked to the test is the consequences of the practice on the validity of the test score interpretation.

Generalizability of score interpretations

Narrowing instructional practices to the methods used in the assessment would not be problematic for the broader interpretation of the construct if it could be shown that higher order thinking skills are generalizable across methods. However, several researchers (Norris, 1989; Perkins & Salomon, 1989) have argued the measurement of critical thinking and problem-solving skills is not generalizable, but is instead context-bound. Thus, it is unclear how generalizable scores are across different forms of assessment. If teaching notebook skills resulted in similar achievement gains across notebooks, direct observations, and other forms of assessment, the construct interpretation would not need to be restricted and the teaching activity remains appropriate under any condition. Limited data are currently available on the generalizability of alternative assessments, but the evidence on the narrower forms of assessment used in paper and pencil tests suggests that generalizability across forms of assessment is limited. Whether examining differences across tests of similar content from different publishers (Koretz, Linn, Dunbar & Shepard, 1991), different item formats (Flexer, 1991), or specific item content and format (Darling-Hammond & Wise, 1985), substantial differences in performance have been found.

In addition, the empirical studies available on the generalizability of alternative assessments suggest that generalizability across tasks has been, and will be, problematic (Dunbar, Koretz, & Hoover, 1991; Linn, 1993). In general, results of alternative assessments show a higher degree of task specific variance even when using the same form or method of measurement than the more frequently reported scorer variance. As Linn (1993) points out, the variance due to raters "can be kept at levels substantially smaller than other sources of error variance, the most notable of which is topic or task specificity" (p. 10). In Baxter et al. (1992), the shared proportion of variance for notebooks and direct observations (ratings on the same hands-on science task) was 0.69 (r=.83). This shared variance suggests that the common construct measured by the two tasks accounts for about two-thirds of the variability in the test scores, whereas about one-third of the variability in test scores is directly attributed to the method of the assessment and other sources of error. In this study, teachers were presumably not teaching to any one form of the assessment. Teaching to one form of the assessment could then have the effect of creating greater method variance, lowering the correlation between the direct observations and the notebooks. While this estimate of method variance is only reported for one study, this type of study needs to be replicated in different content areas and with different methods of assessment to understand more fully the effect of method on authentic test scores.

With the limited generalizability of alternative assessments across tasks, teaching to the particular test form and content (topic) can potentially have large effects on the generalizability of the interpretation that can be made from a test score. Assuming that the test score effectively measures the construct (which may not be possible given the above evidence of generalizability without multiple items or tasks); narrowly teaching to the test would limit the interpretation of the test scores in many situations, particularly for accountability assessments.

Conclusions

Authentic assessment will continue to serve a useful function in low stakes testing (e.g., classroom assessment). However, authentic assessment will probably not solve the problems inherent with high-stakes assessment, and with the variation in methods of assessment, may exacerbate the problem. The nature of high-stakes assessment will continue to be to narrow instruction, which in turn will narrow the interpretation of the construct measured, regardless of the form of assessment. As long as the assessment has high pressure on students, teachers, and administrators to do well, instruction will be narrowed to more accurately reflect the sample of skills, methods and content measured by the assessment, rather than broadened to cover the domain the assessment is intended to sample from and generalize to in the interpretation of test scores.

With authentic assessment, the method of the assessment will lay an integral role given evidence that generalizability across methods is not high. If the method is known in advance of the assessment, instruction will be narrowed to match the assessment method. Instruction would remain broader only if the method of assessment was unknown. However, even if the method of assessment was unknown, the method would still play a large part in the assessment and its interpretation since students who happened to be taught and learn with the particular method used in the assessment would perform better. Thus, the interpretation of the scores would need to be linked to instruction with the particular assessment method.

Ultimately, only two solutions are known for this problem—no high-stakes assessment or multiple high-stakes assessments using multiple methods of assessment. Unfortunately, neither solution will be completely satisfactory. The first solution (no assessment) offers no form of accountability to the public at large nor the parents of the students. The second solution (multiple assessments) will take more time from instruction and learning, as well as increasing the anxiety levels of students, teachers, and administrators. Thus, consistent with paper and pencil assessments in high-stakes situations, authentic assessments may not serve the needs of accountability without narrowing the construct to the particular methods used in the assessment.

REFERENCES

Airasian, P.W. (1988). Measurement driven instruction: A closer look. *Educational Measurement: Issues and Practices, 7* (4), 6-11.

Baxter, G.P., Shavelson, R.J., Goldman, S.R., & Pine, J. (1992). Evaluation of procedure-based scoring for hands-on science achievement. *Journal of Educational Measurement, 29* (1), 1-17.

Darling-Hammond, L., & Wise, A.E. (1985). Beyond standardization: State standards and school improvement. *The Elementary School Journal, 85,* 315-336.

Dunbar, S.B., Koretz, D.M., & Hoover, H.D. (1991). Quality control in the development and use of performance assessments. *Applied Measurement in Education, 4* (4), 289-303.

Flexer, R.J. (1991). *Comparisons of student mathematics performance on standardized and alternative measures in high-stakes contexts.* Paper presented at the annual meeting of the American Educational Research Association and the National Council on Measurement in Education, Chicago.

Koretz, D.M., Linn, R.L., Dunbar, S.B., & Shepard, L.A. (1991). *The effects of high-stakes testing on achievement: Preliminary findings about generalization across tests.* Paper presented at the annual meeting of the American Educational Research Association and the National Council on Measurement in Education, Chicago.

Linn, R.L. (1993). Educational assessment: Expanded expectations and challenges. *Educational Evaluation and Policy Analysis, 15* (1), 1-16.

Linn, R.L. (1983). Testing and instruction: Links and distinctions. *Journal of Educational Measurement, 20* (2), 179-189.

Linn, R.L., Baker, E.L., & Dunbar, S.B. (1991). Complex, performance-based assessment: Expectations and validation criteria. *Educational Researcher, 20* (8), 15-21.

Mehrens, W.A. (1992). Using performance assessment for accountability purposes. *Educational Measurement: Issues and Practice, 11* (1), 3-9, 20.

Mehrens, W.A., & Kaminski, J. (1989). Methods for improving standardized test scores: Fruitful, fruitless or fraudulent? *Educational Measurement: Issues and Practices, 8* (1), 14-22.

Messick, S. (1989). Validity. In R.L. Linn (Ed.), *Educational measurement* (3rd Ed). New York: American Council on Education.

Miller, M.D., & Legg, S.M. (in press). Alternative assessment in a high-stakes environment. *Educational Measurement: Issues and Practice.*

Norris, S.P. (1989). Can we test validity for critical thinking? *Educational Researcher, 18* (9), 21-26.

Perkins, D.N., & Salomon, G. (1989). Are cognitive skills context-bound? *Educational Researcher, 18* (1), 16-25.

Popham, W.J. (1992). A tale of two test-specification strategies. *Educational Measurement: Issues and Practice, 11* (2), 16-17, 22.

Popham, W.J. (1987). The merits of measurement-driven instruction. *Phi Delta Kappan, 68,* 679-682.

Wiggins, G. (1989). Teaching to the (authentic) test. *Educational Leadership, 46* (7), 41-47.

Managing Classroom Assessments: A Computer-Based Solution for Teachers

by

Madhabi Banerji, Ph.D.
Supervisor, Research and Evaluation Services

P. Charles Hutinger
Computer Specialist, Information Services

District School Board of Pasco County
Research and Evaluation Services

Prepared for presentation at the Annual Brievogel Conference of the
Florida Educational Research Council, Gainesville, Florida
March 25, 1993

Managing Classroom Assessments:
A Computer-based Solution for Teachers

Assessment and School Reform: The Larger Picture

The 1990s impetus for public school reform has brought with it a rethinking in the way in which assessment functions are approached and carried out in schools. New understandings of the context in which educational assessment occurs have led to serious thought about the types of assessment programs that should be given priority in schools, and whether classroom assessment should take precedence over externally driven, large scale assessment programs (Stiggins & Conklin, 1992). At issue has also been the kind of information on student learning that assessments should provide (Shepard, 1989; Wiggins, 1989); and whether assessment programs should serve the needs of teachers, students, parents, on the one hand, or policymakers and legislators, on the other (Shepard, 1989; Wiggins, 1989). There has been fairly widespread recognition that tests of multiple-choice formats, particularly those that are externally developed and standardized, cannot provide a complete and balanced picture of student learning in several valued areas of the curriculum. The latter thinking has led to a call for assessments of "alternative" formats. In sum, reforms influencing areas of school organization, curriculum and instruction have significantly impacted how assessment programs are organized and administered in public schools (LeMahieu, 1992).

This paper describes how one school district in west central Florida, the Pasco County School System, is preparing itself to respond to new assessment needs that have arisen as a result of restructuring efforts in elementary schools. Specifically, it presents the description of a computerized resource, designed to assist teachers in linking instructional and assessment management in the classroom, titled the Assessment Management System (AMS). The AMS is a small but important part of a more comprehensive teacher-centered, classroom assessment program that was proposed as a complementary piece to the existing large-scale testing program in Pasco. It provides alternative assessment resources to teachers, as well as tests that are of more traditional, structured-response formats.

Prior to a presentation of the Assessment Management System and specific applications of the same, this paper provides some detail of the rationale underlying and the programmatic context of its development.

Why a New Approach to Assessment was Needed:
The Local Picture

Under the auspices of a Graduation Enhancement Steering Committee, represented by teachers, instructional leaders, administrators and community members, the Pasco County School District began in 1988 to examine ways of restructuring all levels of schooling so as to enhance student learning and graduation rates (see *Graduation Enhancement Steering Committee Proceedings,* District School Board of Pasco County, 1990). The Elementary Task Force, a subgroup of the larger steering committee, recommended that elementary schools be reorganized as nongraded, continuous progress schools, allowing students to move at their own rates through the curriculum (see *School Restructuring for Graduation Enhancement: A Proposal for a Continuous Progress Elementary Program,* District School Board of Pasco County, 1990). As a consequence of the above initiatives, several changes took place in the areas of curriculum and instruction that impacted what students would learn, how they would be taught, how classrooms would be organized, and the new roles teachers would play in the classroom. As new areas were emphasized in the curriculum, a growing discontent was experienced by teachers and instructional leaders regarding the inability of existing forms of assessment to describe what students were learning.

A compelling reason for embarking on a new assessment solution was the need to design assessments that were aligned to the newly valued areas of the district's curriculum, such as an integrated language arts program, a personal and social development curriculum, and programs in math and the content areas that emphasized higher levels of procedural knowledge and problem solving. Schools and teachers were accountable to parents and the larger community for student learning. Existing tests from textbooks and information from the standardized testing program could no longer provide adequate information on the full range of learner outcomes represented in the district's elementary curriculum.

As reforms were implemented at the classroom level, teachers who had been prepared to perform in radically different educational environments found themselves in need of new training, both in instruction and in assessment. Because teachers were encouraged to use flexible and varied instructional strategies, with textbooks serving as resources rather than primary teaching tools, tests and assessments in existing textbook series were often of limited value, underscoring the need for developing customized assessments tied to the local curriculum.

The above changes also made clear that any reforms in the assessment area must include a teacher support system, and concrete, usable assessment resources that teachers could depend upon if they so needed. It was thus concluded that the new arm of the assessment program have a classroom focus, and include a teacher support system incorporating staff development and training.

Specific Program Features and Implications for Assessment

A number of rather fundamental, specific features of the restructured elementary program at Pasco had direct implications on how the assessment program was conceptualized. These features, listed below, influenced the structure and specific attributes of the proposed assessment program.

Feature 1: Holistic View of the Child

Philosophically, the local educational program takes a holistic view of the child, emphasizing development in all areas: cognitive, academic, social, emotional and physical. Students are also respected as unique individuals.

Implications for Assessment: To support the philosophy of the *whole child*, assessment methods needed to address all areas of student development, as opposed to focusing primarily on academic areas, as was typical in traditional testing programs.

Feature 2: Outcome-based Curriculum

The building blocks of the curriculum are ***Outcomes***, defined as descriptions of what students will learn and be able to do when they complete the elementary program (see *Pasco 2001: A Community of Connected Schools*, District School Board of Pasco County, 1992). Specific descriptions of knowledge, concepts and skills subsumed under the outcomes are labeled as ***Indicators***, organized at two approximate levels of complexity, primary and intermediate. Figure 1 presents an example of two Outcomes in mathematics, M1 and M2, with primary and intermediate indicators subsumed. Classroom instruction, delivered in the form of themes or units, is destined to address particular outcomes and indicators.

Outcomes and ***Indicators*** are subject area-specific, and naturally group under traditional disciplines such as mathematics. However, they can also be organized **across** disciplines using interdisciplinary organizers called ***Performance Roles***. The ***Performance Roles*** represent the broad roles in which students should be able to perform when they exit the program, such as "Problem Solver," "Informed Communicator," "Conceptualizer." Because they provide a conceptual mechanism for interrelating outcomes from different subject areas, ***Performance Roles*** serve as an effective means for designing and delivering interdisciplinary instruction. Figure 2 presents an interdisciplinary grouping of outcomes under the Performance Role, "Problem Solver."

Figure 1. Outcomes Organized by Subject Area: Mathematics Examples

MATHEMATICS OUTCOMES

M1 **Solve mathematical problems by applying various problem-solving strategies.**

Indicators: Primary	Select from a number of strategies (such as building models, drawing diagrams, or acting out problem situations) to show the inverse relationship between addition and subtraction.
Intermediate	Select from a number of strategies (such as building models, drawing diagrams, or acting out problem situations) to show the inverse relationship between multiplication and division.

M2 **Communicate using the language and symbols of mathematics across disciplines and situations.**

Indicators: Primary	Read and write number sentences using one or more of the following mathematics symbols: +, -, =, <, >.
Intermediate	Read and write number sentences using one or more of the following mathematics symbols: +, -, <, =, x, <, >, ≤, ≥, %,)—, +, π, =, (exponents), (fraction symbols).

Figure 2. Outcomes in Three Areas Organized by Performance Role: Problem Solver

PERFORMANCE ROLE: PROBLEM SOLVER

Math 3	Applying reasoning to justify solutions, teaching processes and conjectures when solving mathematical problems.
Communication 6:	Self-monitor and correct own performance when reading/writing/speaking/listening.
Social Development 8:	Use problem-solving strategies to make good personal decisions and choices for behavior.

The concept of ***Performance roles*** is consistent with recent ideas in educational reform that define exiting student outcomes as broad, generic competencies that cut across subject area borders and require culminating demonstrations of learning by students upon completion of the program (Spady, 1992; Redding, 1992).

Implications for Assessment: A direct implication of an outcome-based curriculum structure on assessment is that assessments be aligned to prioritized outcomes and indicators in particular phases of the instructional cycle, an approach that is consistent with accepted standards for the design of sound classroom tests (Nitko, 1989).

In order to assess students on broad, culminating ***Performance Roles*** that cut across subject areas and required student demonstrations of complex performances or the development of student products, new assessment possibilities had to be explored. Test items focusing on singular skills and/or concepts were no longer valid representations of complex sets of learning outcomes. Assessment exercises now needed to represent combinations of outcomes and indicators that students were called upon to apply in the context of completion a multistep problem or task.

Several subject-centered outcomes, such as the application of reading strategies during oral reading, focused on behaviors that did not lend themselves to assessment by written, structured-response tests. Depending on the outcome(s) to be assessed, a variety of assessment methods, such as product assessments, process assessments, and written tests, were needed to assist teachers in making necessary decisions on student learning.

Feature 3: Nongraded School Organization

Continuous progress elementary schools are reorganized into primary and intermediate houses consisting of multiage clusters of students served by teams of teachers who stay with the same children for three consecutive years of schooling. In the continuous progress organization, traditional grade levels do not exist, and students move at their own pace through the curriculum.

Implications for Assessment. Traditional assumptions about groups of same-age students learning the same body of knowledge and skills, within fixed year-long blocks of time (grade levels), are no longer true in continuous progress schools. In a continuous progress context, common assessments for large groups of students are not as critical as timely assessments tailored to the needs of individuals or small groups of students. To accommodate variable assessment needs, the assessment program had to provide teachers with flexible options in both the selection and administration of assessments.

Feature 4: Redefined Roles of the Teacher.

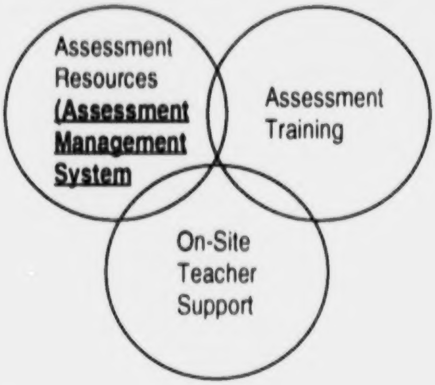
In the restructured house, teachers work in teams to meet student needs. They collaborate as designers, facilitators, and decision-makers in the classroom in every area impacting the teaching and learning process.

Implications for Assessment. The role of teachers as assessment designers and managers is directly impacted within the continuous progress team context. To prevent the teachers' task from becoming unduly burdensome, tools and resources are needed that can assist the efficient monitoring and assessment of the learning process. Management systems that assist teachers in conducting routine classroom functions, such as instruction, assessment, reporting and record keeping, should relieve them of mundane work, allowing more time and freedom for creative instruction. Systems that assisted teachers to integrate instructional and assessment tasks should not dictate what teachers do, but rather, they should provide teachers with multiple options and information that enlighten classroom decisions and help reduce their work load.

The Proposed Solution: Specific Components of Classroom Assessment Program to Support Continuous Progress

A three-pronged solution was taken in designing the new arm of the assessment program to support local changes occurring in the areas of curriculum, instruction, and school organization. The solution attempted to take into account the redefinition of teacher roles in the classroom. The three-part solution, illustrated diagrammatically in Figure 3, consisted of overlapping components that included: (1) the development of a classroom assessment resource for teachers in the form of a computerized Assessment Management System (AMS) which enabled a linking of instructional and assessment decision making; (2) the provision of optional but formal teacher training in assessment methodology to support teachers in their expanded role as assessors and decision makers in the classroom; and (3) the development of an "assessment leader" at the school site to provide on-site guidance and support to teacher in their day-to-day tasks as instructional and assessment managers.

Figure 3. Components of Classroom Assessment Program



The Assessment Management System (AMS): A Prototype

This paper now shifts its focus to the first part of the three-pronged solution being implemented at Pasco, namely, the Assessment Management System (AMS).

Despite its title focusing on the "assessment" end of classroom functions, the Assessment Management System (AMS) is really intended to help teachers manage all aspects of the complete instructional design process, i.e., deciding what to teach, designing and delivering instruction, deciding how and when to assess students, selecting appropriate assessments, designing assessments, making decisions on student learning and/or evaluating instruction (Tyler, 1949; Dick and Carey, 1971).

The AMS, at this stage a prototype, consists of a computerized system that allows teachers to link instructional planning with assessment planning in the beginning of a term or semester, starting with ***Outcomes*** or ***Performance Roles*** as the initial organizer.

In developing the AMS as a resource for teachers, three questions were asked: (1) What assessment design questions is a teacher, or a team of teachers, likely to face in the beginning of a semester or quarter in a Continuous Progress elementary classroom? (2) How can a system be developed so as to reduce the assessment design work of teachers to one of judicious selection, or to one of adaptation of already existing assessments to specific teaching-learning contexts? and (3) How can a system be designed to promote the interlinking of instructional and assessment functions, so that teachers can make smooth transitions between both domains of decision making?

To address the above questions, the first version of the AMS took the form of a set of interactive databases and files that would allow teachers to efficiently search through, find, edit, or print assessments tied to particular curriculum ***Outcomes*** and ***Performance Roles***. For each outcome, multiple assessment tools are available, in most cases requiring different response-formats from students—such as written, open ended; written, structured response; behavior assessments; product assessments—providing an array of options to teachers.

The chart shown in Table 1 illustrates the different options that the current version of the AMS provides to teachers for conducting various classroom tasks. For instance, the AMS could assist teachers in reviewing and selecting curriculum outcomes upon which to focus instruction for a semester or quarter. Using the AMS' search and find capabilities, teachers could view an interdisciplinary set of outcomes that are tied to a given Performance Role, such as Conceptualizer. Such combinations of outcomes could spark ideas for the design of an interdisciplinary unit of instruction. Once outcomes are selected for instruction, teachers could examine assessment tasks and exercises in the AMS that were specifically designed to assess particular outcomes at the primary and intermediate level. As indicated earlier, Assessment Methods in the AMS are assessments of alternative and traditional formats, with attached scoring systems, that teachers can print directly if they so desire.

Table 1. What can the Assessment Management System (AMS) do?

Task	Typical Questions asked by teachers:	How the AMS can help:	Sample screens
Instructional Planning for semester/quarter	What outcomes and indicators will we target for instruction?	Look up outcomes and indicators in given subject areas	Figure 6
Instructional Planning for Interdisciplinary Unit or Themes	Which outcomes lend themselves to teaching an interdisciplinary theme?	Look up outcomes across subject areas linked by Performance Roles	Figure 6
Assessment Selection	Are there assessments available that I can adapt or select to fit my specific teaching learning situation?	Look up Assessment Methods tied to outcomes	Figure 7
Assessment Evaluation/ Adaptation	Is this assessment suitable for my context and purposes?	Look up Assessment Specifications	Figure 7
Assessment Design	Can I edit or redesign assessments?	Look up Assessment Specifications; Add new assessments to AMS.	Figure 8

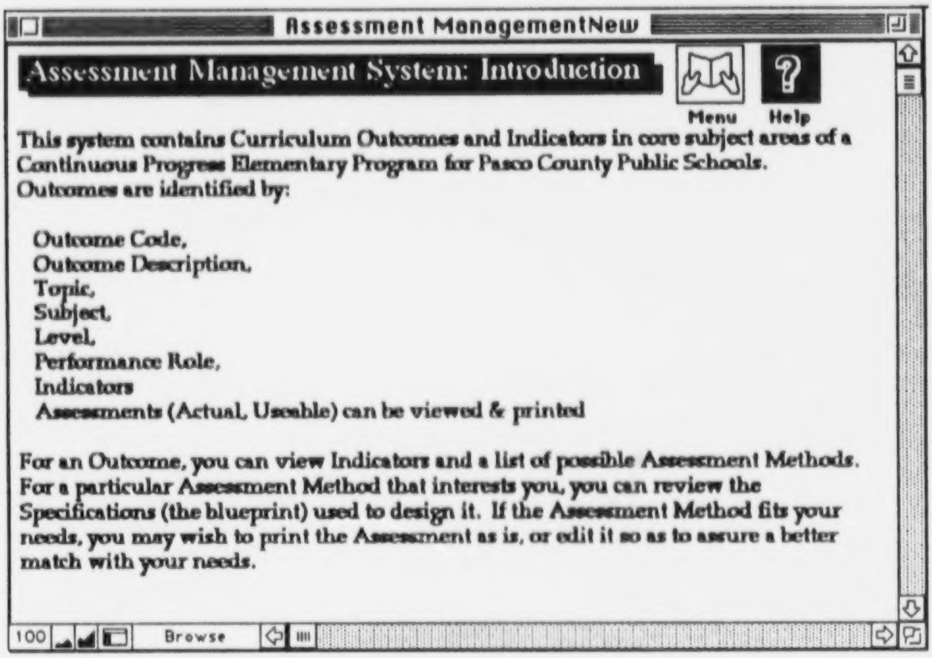
The AMS prototype provides three to four optional Assessment Methods for each outcome, and also allows teachers to examine the specifications used to design the assessments. Through a review of the design

BEST COPY AVAILABLE

specifications, teachers may evaluate the extent of alignment between their teaching-learning context and particular assessment exercises.

The AMS is menu-driven. Figures 4-8 present sample screens from the AMS, that help accomplish some of the tasks described. In Figures 4-5, the teacher is introduced to the purposes and content of the AMS and oriented to using the system with a "Help" screen. Figure 6 illustrates a screen with an outcome listing that would appear if teachers chose the "View Outcomes" option from the menu. Figures 7 and 8 show screens that would appear if a teacher were interested in outcome C7.0, on Language Conventions, and wanted to view available Assessment Methods for the same. The handwriting scale, illustrated in Figure 8, is included among assessment resources that accompany a primary level language arts portfolio in the AMS.

Figure 4.



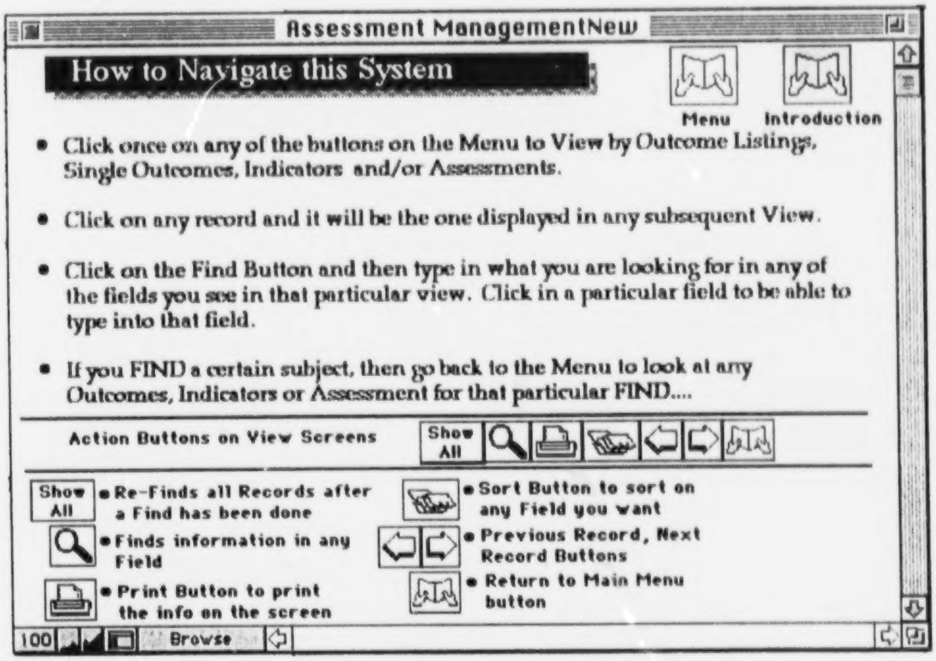
Assessment Management System: Introduction

This system contains Curriculum Outcomes and Indicators in core subject areas of a Continuous Progress Elementary Program for Pasco County Public Schools. Outcomes are identified by:

Outcome Code,
Outcome Description,
Topic,
Subject,
Level,
Performance Role,
Indicators
Assessments (Actual, Useable) can be viewed & printed

For an Outcome, you can view Indicators and a list of possible Assessment Methods. For a particular Assessment Method that interests you, you can review the Specifications (the blueprint) used to design it. If the Assessment Method fits your needs, you may wish to print the Assessment as is, or edit it so as to assure a better match with your needs.

Figure 5.



How to Navigate this System

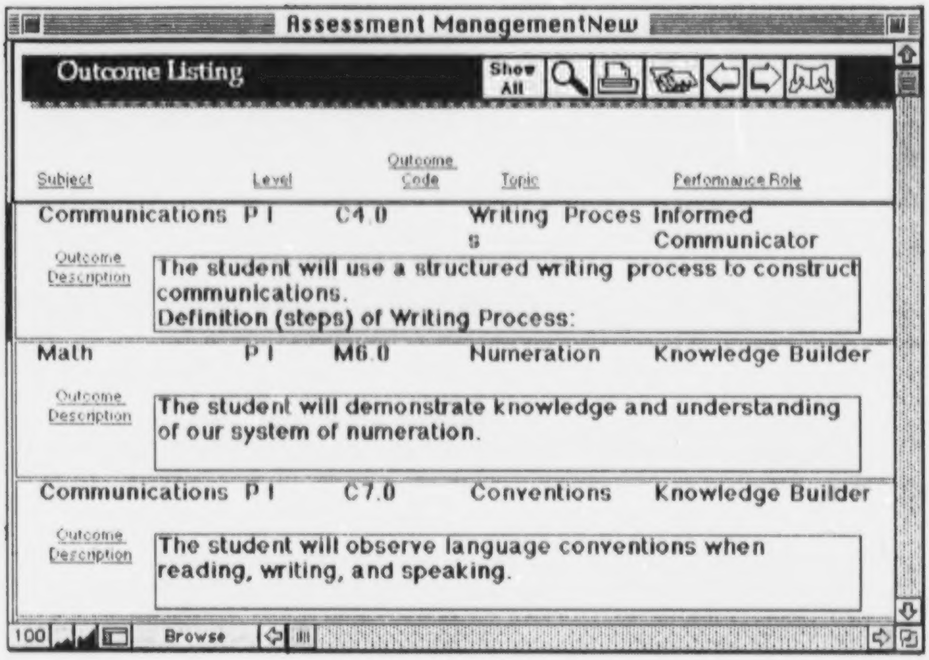
- Click once on any of the buttons on the Menu to View by Outcome Listings, Single Outcomes, Indicators and/or Assessments.
- Click on any record and it will be the one displayed in any subsequent View.
- Click on the Find Button and then type in what you are looking for in any of the fields you see in that particular view. Click in a particular field to be able to type into that field.
- If you FIND a certain subject, then go back to the Menu to look at any Outcomes, Indicators or Assessment for that particular FIND....

Action Buttons on View Screens

Show All	Re-Finds all Records after a Find has been done
🔍	Finds information in any Field
🖶	Print Button to print the info on the screen
	Sort Button to sort on any Field you want
⇦ ⇨	Previous Record, Next Record Buttons
	Return to Main Menu button

BEST COPY AVAILABLE

Figure 6.



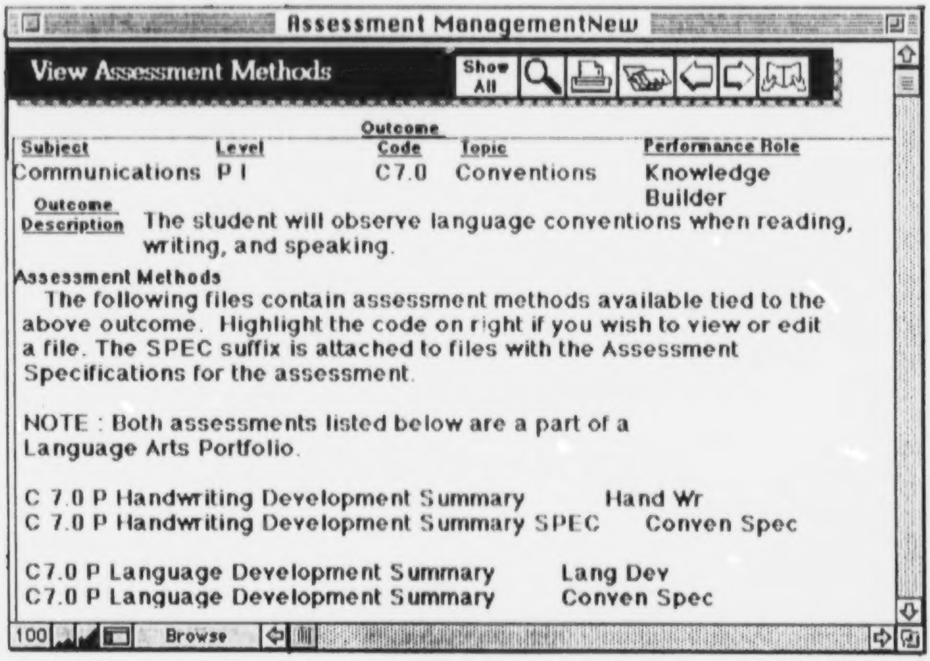
Subject	Level	Outcome Code	Topic	Performance Role
Communications	P I	C4.0	Writing Process	Informed Communicator
Math	P I	M6.0	Numeration	Knowledge Builder
Communications	P I	C7.0	Conventions	Knowledge Builder

Communications C4.0 – Outcome Description: The student will use a structured writing process to construct communications.
Definition (steps) of Writing Process:

Math M6.0 – Outcome Description: The student will demonstrate knowledge and understanding of our system of numeration.

Communications C7.0 – Outcome Description: The student will observe language conventions when reading, writing, and speaking.

Figure 7.



Subject	Level	Outcome Code	Topic	Performance Role
Communications	P I	C7.0	Conventions	Knowledge Builder

Outcome Description The student will observe language conventions when reading, writing, and speaking.

Assessment Methods

The following files contain assessment methods available tied to the above outcome. Highlight the code on right if you wish to view or edit a file. The SPEC suffix is attached to files with the Assessment Specifications for the assessment.

NOTE : Both assessments listed below are a part of a Language Arts Portfolio.

C 7.0 P Handwriting Development Summary — Hand Wr
C 7.0 P Handwriting Development Summary SPEC — Conven Spec

C7.0 P Language Development Summary — Lang Dev
C7.0 P Language Development Summary — Conven Spec

BEST COPY AVAILABLE

Figure 8. Sample Assessment for Communications Outcome C7.0: The student will observe language conventions when reading, writing, and speaking.

Handwriting Development Summary

Name: ______________________________ Grade: ____________

Code:
✓ = Consistent Evidence of this E = Early Signs N = No Evidence Yet U = Unable to determine

Sample Description (Letter, Manuscript, Practice)

1. ______________________ 4. ______________________
2. ______________________ 5. ______________________
3. ______________________ 6. ______________________

Date	1st year		2nd year		3rd year	
1. Holds pencil properly						
2. Shows left to right directionality						
3. Has correct posture when writing						
4. Has correct paper position when writing						
5. Identifies upper case letters (How many?)						
6. Forms upper case letters legibly (How many?)						
7. Identifies lower case letters (How many?)						
8. Prints lower case letters legibly (How many?)						
9. Spaces correctly between letters in a word						
10. Spaces correctly between words in a sentence						
11. Demonstrates mastery of upper case alphabet in writing						
12. Demonstrates mastery of lower case alphabet in writing						
13. Prints legibly						
14. Writes legibly in cursive						
15. Other indicators: Neatness						

Please fill in this section based on discussion with student and parents.

What can we work on next? ______________________________

Note to the Teacher:

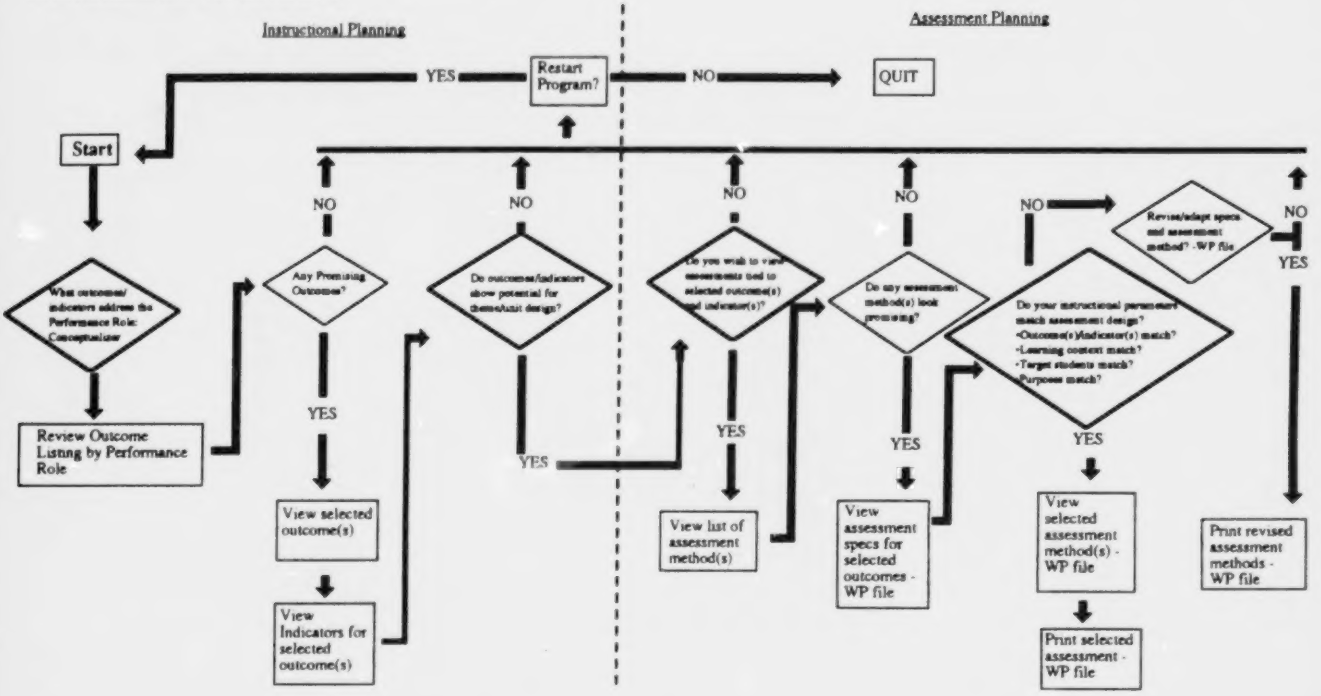
This instrument must be filled by the teacher based on observations made of the child in class as well as examination of handwriting samples produced. Share with parents at appropriate times.

As suggested previously, teacher training on the AMS is a necessary and important part of the implementation plan of the new assessment solution. Such training will prepare teachers in utilizing the software and hardware, as well as in systematic decision-making processes related to teaching and learning. How can the AMS be used by teachers to make decisions that cross the instructional and assessment domains? Figure 9 shows in flow chart form, an application for an interdisciplinary unit. Starting with the Performance Role, Conceptualizer, teachers may search through the database for outcomes in different subject areas that will help students become proficient conceptualizers. A systematic search through the database will lead teachers to assessments linked to "Conceptualizer" outcomes, and specifications used to design them, finally bringing them to a decision-making point where they could either print the available assessment or adapt/redesign it.

Figure 9.

ASSESSMENT MANAGEMENT SYSTEM - APPLICATION

Planning instruction and assessment for an interdisciplinary theme designed for the Performance Role: CONCEPTUALIZER


Instructional Planning
Assessment Planning
Restart Program?
YES
NO
QUIT
Start
What outcomes/indicators address the Performance Role: Conceptualizer
Review Outcome Listing by Performance Role
Any Promising Outcomes?
NO
YES
View selected outcome(s)
View Indicators for selected outcome(s)
Do outcomes/indicators show potential for theme/unit design?
NO
YES
Do you wish to view assessments tied to selected outcome(s) and indicator(s)?
NO
YES
View list of assessment method(s)
Do any assessment method(s) look promising?
NO
YES
View assessment specs for selected outcomes - WP file
Do your instructional parameters match assessment design?
•Outcome(s)/Indicator(s) match?
•Learning content match?
•Target students match?
•Purpose match?
NO
YES
View selected assessment method(s) - WP file
Print selected assessment - WP file
Revise/adapt specs and assessment method? - WP file
NO
YES
Print revised assessment methods - WP file

BEST COPY AVAILABLE

AMS Design Issues: Hardware and Software Needs

The prototype of the AMS was built on an Apple Macintosh platform and requires the following hardware and software capabilities:

Apple Macintosh Computer (preferably LC class or better)
System 7 Operating System
FileMaker Pro 2.0
Word Processing Software (Microsoft Works Version 2.0 was used)
A Hard drive, preferable 80 mega bytes or more
At least 4 mega bytes of RAM, preferably 10 (to accommodate multiple applications to be open at the same time).
A Laser Printer

The design of the prototype database was driven by what the teachers needed to be able to do—namely, look up or find information by any field name; view or browse through any of the fields; search and list records by field, such as performance role, level, subject; actually pull up an assessment to print and use, or review specifications.

To meet the initial software capabilities needed to store and retrieve outcomes, indicators, and assessment tools, FileMaker Pro by Claris was selected to serve basic database needs. This software is a powerful database that allows the design of flexible and simple front ends to the data for the users. Once the concepts of assessment became concrete, eight critical fields were identified that contributed to the core structure of the database. These were Outcome Code, Outcome Description, Topic, Subject, Level (primary, intermediate), Performance Role, Indicators and Assessment Methods.

Actual Assessments and Assessment Specifications were in the form of word processing files (WP files) that could be retrieved while in FileMaker Pro, using a File Find capability of the System 7 operating system. Irrespective of the WP software program that produced the assessment, the file could be automatically launched for viewing, editing or printing. The database, in the meantime, was still available for searches and lookups.

When using the software, teachers also needed the capability of adding to the database, namely, creating their own assessments. WP files were more user friendly and easy to access, edit and print than the relatively complicated, structured database of FileMaker Pro. It was thus decided that teachers would be more comfortable using word processors with which they were familiar, and Microsoft Works was chosen to serve this purpose.

Other data fields, such as Outcome Description, Indicators, Performance Role, were kept unalterable within the FileMaker Pro structure, to maintain the integrity of the system. These fields were designed to be selected from a list of choices, thereby prohibiting free typing and insuring correct spelling. In this way, all searches done of these two fields would not be hindered by potential typographical errors.

Teacher Training Issues

To facilitate comfortable use of the system by teachers, future training has to address at least three areas:

1. Knowledge of what makes technically sound classroom assessments;
2. Proficiency in using the AMS software and hardware; and
3. Easy integration of the AMS into teachers' regular classroom activities, so that the technology helps make classroom functions more efficient rather than burdensome.

All three areas are being considered very seriously as the district moves forward from the prototype to a more refined version of the AMS. An introductory teacher training institute has been developed titled "Technology Links for School Improvement" (DSBPC, 1992) that touches upon technical concepts related to classroom assessment design and provides a demonstration of the AMS. Supplementary training programs are in preparation to support teachers in making the transition to their new role in the Continuous Progress Classroom.

Future Development

As mentioned earlier, the AMS is the first version of the concept that is hoped will evolve into a more comprehensive and refined instructional management system for teachers. Future components of the system may include a database of student records where scored assessments can reside, with the possible capability of integrating information relevant to all aspects of the teaching-learning process.

The quality and effectiveness of the AMS will depend, in large part, upon the quality of the assessments that are housed in it. A full description of the formal process of assessment design is beyond the scope of this paper. However, it will be mentioned that a technically rigorous approach is being taken in developing assessments for the redesigned elementary curriculum. The design process includes validation and reviews by measurement and curriculum experts, as well as field testing of products that will enable an examination of their psychometric properties. Although the AMS will not contain an exhaustive supply of assessments, it is intended that it hold well-designed assessments that teachers can use as models. As teachers get more proficient in formal assessment design, products developed by them could eventually be reviewed for technical quality and considered for possible inclusion in the AMS.

The immediately following version of that system will contain assessments designed for the district's elementary mathematics curriculum. According to the district's plan, the AMS will then be ready for a pilot in primary and intermediate houses. Program evaluation information from the first tryout will be used for future development work. Development of the AMS is expected to continue over the next two years.

REFERENCES

Dick, W. & Carey, L.M. (1978). *The systematic design of instruction.* Glenview, IL: Scott, Foresman & Co.

District School Board of Pasco County (1990). *Graduation Enhancement Committee: Proceedings.* Land O' Lakes, FL: Author.

District School Board of Pasco County (1990). *School Restructuring for Graduation Enhancement: A Proposal for a Continuous Progress Elementary Program.* Land O'Lakes, FL: Author.

District School Board of Pasco County (1992). *Pasco 2001: A Community of Connected Schools.* Land O' Lakes, FL: Author.

District School Board of Pasco County (1992). *Technology links for school improvement: Curriculum, instruction and assessment* Land O' Lakes, FL: Author (In collaboration with the Bureau of Teacher Education, Florida Department of Education).

LeMaheiu, P.G. (1992). Defining, developing a well-crafted assessment program. *NASSP Bulletin.* 76, 545, 50-56.

Nitko, A.J. (1989). Designing tests that are integrated with instruction. In R.L. Linn (Editor), *Educational Measurement.* New York, NY: American Council on Education and the Macmillan Publishing Co.

Shepard, L.A. (1989). Why we need better assessments. *Educational Leadership.* 46, 7, 4-9.

Spady, W. (1992). It's time to take a close look at outcome-based education. *NASP Communique.* May Issue.

Stiggins, R.J. & Conklin, N.F. (1992). *In teachers' hands: Investigating the practices of classroom assessment.* New York: SUNY.

Redding, N. (1992). Assessing the big outcomes. *Educational Leadership.* May Issue.

Tyler, R.W. (1949). *Basic principles of curriculum and instruction.* Chicago, IL: University of Chicago Press.

Wiggins, G. (1989). Teaching to the authentic test. *Educational Leadership.* 46, 7, 41-47.

APPENDICES

SAMPLE ASSESSMENT MATERIALS FROM THE ASSESSMENT MANAGEMENT SYSTEM

Appendix A

Code: Weather P2

Title of Assessment Method:
Group Interview/Discussion on Weather Concepts and the Scientific Method

Form A: Teacher Prompt Sheet

Note: These questions will help to set up the context to enable observations of oral communication behaviors and student understanding of the scientific method and weather terms. Use the attached recording form to document performance of individual children (Form B). Please note that these are suggested questions that could be adapted or modified to fit particular teaching-learning situations.

Today, we will talk about the most important things that we learned in the Weather Theme.

Let me begin by asking you:
1. *What did you learn that was NEW?*

Why was it new to you?
(Call names as needed)

What did you enjoy learning the most? Making: Weather Calendar?

Let's take a look at your Weather Calendar.

Scientific Method

2. *What question(s) were you trying to answer when you began keeping a daily weather calendar?*

*2a. *How did the Weather Calendar help you understand weather? (call names as needed)*

3. *Johnny, let's take a look at Monday the 15th (select appropriate day) on your calendar. Why did you choose a "sunny" picture for this day?*

4. *Who else had a "sunny" picture on the same day? Do we have a consensus?*

4a. *Did anyone have a different picture on that day? Why did you choose that one?*

5. *What did this column (point to estimate column) help you do? (call names as needed)*

5a. *When scientists guess what they expect the results of a scientific study to be, what word do they use?*

6. *Where did you record what you observed? (call names as needed)*

*7. *Compare what you guessed with what you recorded on ______ day. (call names as needed)*

*7a. *Compare what you guessed with what you observed for the whole month. (call names as needed)*

**8. *What do you think next week's weather will be like, based on what you recorded this week? What makes you say that?*

**9. *What do you think the weather will be like next May based on your study of May weather this year?*

***9a. *If you were to do this study again next year, would your guesses for rainy, sunny and cloudy days be the same?*

***9b. *What would you do differently?*

***10. *Now that you've learned about how you can study and predict natural events like weather, where else would you use this method? How will you use it?*

Weather Facts and Terms

11. *Can you think of the word we used that describes when water changes its state? (Evaporation, water cycle)*
    Who remembers?
    Do you remember the experiment we did?
    What did we learn?

Recording and Interpreting Data

12. *Point out the legend on your calendar. How did the legend help you record what you saw?*

13. *If you had a day on which it looked like it would rain but didn't, how would you show it on your calendar?*

14. *How many cloudy days are recorded in your calendar?*

15. *How many sunny days are recorded?*

16. *What do your records show—more rainy days or more sunny days? How many more?*

Appendix B

Code: Weather P2

Title of Assessment Method: Group Interview/Discussion on Weather Concepts and the Scientific Method

Form B: Student Recording Form

Student Name: ______________________________

[To be used in conjunction with Form A]
Check if observed during Discussion/Interview

Level: ______________

I. Oral Communications Behaviors

Responds to teacher questions ☐
Responds to student questions ☐
Attends/Listens to others ☐
Introduces new ideas ☐
Builds on other's ideas ☐
Poses Questions ☐
Other Behaviors (Describe) ☐

II. Knowledge/Understanding of Weather Facts/Vocabulary

Uses terms in correct contexts ☐

"Listen for" (Circle all that were used)

sunny	windy
cloud types	water cycle
evaporation	hazardous
weather	weather instruments

Other words: ____

III. Knowledge/Understanding of Scientific Method

- Interprets data correctly on calendar ☐
- Reads calendar correctly ☐
- Reads legend correctly ☐

Appendix B (continued)

- Distinguishes between estimates (hypothesis) and observations ☐
- Makes reasonable predictions ☐
- Attempts transfer of knowledge ☐

Evidence of extended vocabulary: ______________________________

Evidence of extended concept learning: ______________________________

Misconceptions identified: ______________________________

Comments: ______________________________

____

Appendix C

Code: ***Weather P2***

ASSESSMENT SPECIFICATION

Title of Assessment Method: *Group Interview/Discussion on Weather Concepts and the Scientific Method*

The assessment was designed for *(check):*

☐ A single discipline unit √ An interdisciplinary theme ☐ Other (explain)

Theme/Unit Title: *"What About Weather"*

Suggested Level *(check):* √ Primary ☐ Intermediate

Outcomes and Indicator(s) to be assessed:

Outcome code:	Indicator(s)
C2	*Use an oral communication process to demonstrate understanding of the scientific method and selected weather facts and vocabulary*
C15	*Use legends and symbols to depict weather patterns in a weather calendar*

Selected Assessment method:

Structured, Oral Discussion/Interview to assess:

(1) Facility in oral communication
(2) Understanding of factual knowledge presented in theme

Description of assessment method:

Purpose for assessment:
Summative - to be administered at the end of 4-6 week theme with reference to weather calendar and other products developed (fact book, weather report) during theme

Characteristics of Target Students:
Emergent, Beginning and Transitional Learners

Learning Context:
Learning opportunities included daily weather observations and experiments (e.g., on evaporation); instruction in the scientific method as it applies to the Weather theme; and specific uses of weather vocabulary. In a large group setting, the teacher used a KWL framework to set directions for instruction and get baseline information on the entry-level knowledge of learners. Periodic discussions, in various grouping formats, were held to verify what was learned and to clarify misconceptions. Fact books were maintained by students to summarize their learning. Students were provided with opportunities to share new learning.

Type of scale or instrument:

A list of teacher prompts (Form A) are accompanied by an observation checklist for individual students (Form B)

Criteria for judging (S)uccessful and (E)xtended performance:

Emergents:	*Accurate responses to questions involving comprehension and comparison (marked with single asterisk - *)*
Beginners:	*Accurate responses to questions involving prediction (**)*
Transitional:	*Accurate and reasonable responses to questions involving transfer of knowledge (***)*

Assessment Conditions

This assessment is a culminating activity for the theme and ties in closely with other product assessments (e.g., Weather Calendar) for the theme. Students may be allowed to look at their products as they participate in this assessment exercise.

The assessment is conducted in a structured small group setting (3-4 students with a teacher and observer). The assessment focuses on three major areas:

1. *Oral communication behaviors in small groups*
2. *Knowledge/understanding of weather facts/and vocabulary*
3. *Knowledge/understanding of the scientific method*

A series of prompts are provided (see Form A: Teacher Prompt sheet) to set up the group exercise and elicit responses in each of the above areas from students. An observation checklist (see Form B) is present to record individual student behaviors and responses to specific questions presented by the teacher and to others in the group. It is suggested that checking/recording be done by an independent observer, while the teacher conducts/guides the oral discussion. An alternative may be that the session is videotaped while the teacher conducts the discussion. Checking of individual student performance may then be conducted at a later viewing of the taped session.

Appendix D

Links Participant Materials
Session 9.0

ILLUSTRATION

Code: CHILD L 1.0, 2.0, 3.0
M. Banerji & J. Sumner, 1992

Language Arts Portfolio

To the Teacher & Parent(s):

Purpose:

The contents of this portfolio are designed to assess the following goals for grade levels K-2 in Project CHILD.

L1.0 Compose descriptive, narrative and practical pieces in writing.
L2.0 Edit using language conventions of punctuation, sentence structure, capitalization, and paragraphing.
L3.0 Demonstrate manuscript handwriting skills.

Portfolio Contents

1. Writing Samples
   At least four writing samples, two from each quarter (4 per semester), must be included, with at least **one** sample in each of these categories:
   - Descriptive Writing (e.g., personal experience)
   - Narrative Writing (e.g., a story)
   - Practical Writing (e.g., a letter)

   Samples may include final drafts or works in progress. Samples will be selected by student with teacher participation as best examples of his/her work for that semester. Samples may be drawn from journal entries, original stories, retold stories, letters, poems or rhymes.

2. Language Expression Development Summary (Attachment 1)
   Checked by teacher once every semester with comments. This is a summary record.

3. Handwriting Samples - Four per semester
   These may be the same as the Writing Samples.

4. Handwriting Development Summary (Attachment 2)
   Checked by teacher once every semester. This is a summary record.

5. Cover Letter Written by Student at the End of School Year
   This letter will be the student's reflection and self-evaluation of what he/she has learned during the past year. In it, the student will justify why the pieces in the portfolio were chosen as the best pieces of his/her work for that year.

Suggestions:
Kindergarten - 2nd Grade Students: The letter may be written to parent(s) or to the new teacher of the following grade. It could be used as a goal-setting activity for individual children for the following year.

Appendix E

District School Board of Pasco County
Project CHILD
Code: CHILD L 2.0,3.0
M. BANERJI & J. SUMNER, 1992

Language Expression Development Summary

Based on at least four selected writing samples every semester.

Name: ______________________________ Grade: __________

	Sample Description (Letter, Story, Journal)	Date Reviewed	Teacher Comments
1.			
2.			

Number of samples produced on Word Processor =

Codes:

✓ = Consistent Evidence of this E = Early Signs of this N = No Evidence Yet U = Unable to to Determine

	Date produced	*Semester 1*	*Semester 2*
A. Ideas Expressed Mainly Through Pictures			
1. Mimics writing - Scribbles			
2. Draws pictures to tell story			
3. Labels pictures to express ideas			
4. Gives oral explanation of pictures			
5. Organizes ideas in sequence when telling story			
6. Writes name			
B. Ideas Expressed Mainly Through Written Words			
7. Relates print to picture			
8. Writes complete words or phrases to describe pictures			
9. Uses invented spelling How many? (Optional)			
10. Uses conventional spellings How many? (Optional)			
11. Uses vocabulary learned in other contexts (e.g., Reading Class) How many?			
12. Gives oral description of Written Word(s) Phrases			

Appendix E (continued)

C. Ideas Expressed Mainly in Sentences and Paragraphs	*Date produced*	*Semester 1*	*Semester 2*
13. Expresses ideas in complete sentences			
14. Sequences sentences logically			
15. Punctuates correctly			
• uses periods correctly			
• uses question marks correctly			
• uses exclamation marks correctly			
• uses capitalization correctly			
• other (explain) ______________________			
20. Uses variety in sentence structures (interrogative, declarative)			
21. Writes paragraphs with beginning, middle, & end.			
• introductory sentence(s)			
• supporting sentence(s)			
• concluding sentence(s)			
22. Uses expansive vocabulary			
23. Other indicators of Language Development:			
Thoughtful ideas			
Original ideas			
Lengthy narratives			
______________________			
______________________			
______________________			

Please fill this section based on discussion with student and parents.

What can we work on next? ____

____

Appendix F

District School Board of Pasco County Project CHILD Code: CHILD L 2.0,3.0
M. Banerji & J. Sumner, 1992

Handwriting Development Summary

Name: ______________________________ Grade: __________

Codes:
✓ = Consistent Evidence of this E = Early Signs of this N = No Evidence Yet U = Unable to Determine

	Sample Description (Letter, Manuscript, Practice)	Date Reviewed	Teacher Comments
1.			
2.			

Date produced	*Semester 1*	*Semester 2*
1. Holds pencil properly		
2. Shows left to right directionality		
3. Has correct posture when writing		
4. Has correct paper position when writing		
5. Identifies uppercase letters (How many?)		
6. Forms uppercase letters correctly (How many?)		
7. Identifies lowercase letters (How many?)		
8. Prints lowercase letters correctly (How many?)		
9. Spaces correctly between words in a sentence		
11. Demonstrates mastery of uppercase alphabet in writing		
12. Demonstrates mastery of lowercase alphabet in writing		
13. Prints legibly		
14. Writes legibly in cursive		
15. Other indicators: Neatness		

Please fill in this section based on discussion with student and parents.

What can we work on next? ______________________________

Note to the Teacher:
This instrument must be filled by the teacher based on observations made of the child in class as well as examination of handwriting samples produced. Share with parents at appropriate times.

Historical Roots of Current Practice in Educational Testing

by

Annie W. Ward and Mildred Murray-Ward
The Techne Group, Inc.
Daytona Beach, FL 32118

HISTORICAL ROOTS OF CURRENT PRACTICE IN EDUCATIONAL TESTING

INTRODUCTION

In the course of the history of educational measurement, a body of knowledge and theory has been developed to guide practice. This paper presents a very brief overview of major events and people that have influenced assessment as educators know it today. Table 1 presents a brief overview of this history. For each period, testing practices are described and the issues involved are discussed. In addition, interested readers may wish to read the original sources listed in the bibliography to learn more about this important part of the "culture of education."

The history of testing spans more than 5,000 years, during which time tests have undergone and will continue to undergo many changes. Many "new" testing procedures currently being advocated have already been tried numerous times, and problems associated with them are well-known, although some of the advocates of new practices tend to ignore the problems.

If we ignore what is known about these problems, failed solutions of the past may be repeated when they are presented as "new movements" in assessment.

ANCIENT BEGINNINGS

Civil Service Testing in China

Testing was a part of the culture of several ancient peoples. Dubois (1964, 1970) provides a description of a very early and long lasting testing program that began in China more than three thousand years ago and lasted until this century. At the time that the testing was started, the Chinese did not have a hereditary aristocracy to provide a governing class nor did they have a university system; therefore, they developed an elaborate examination system to select, promote, and retain public officials. The intent was that government officials would be well-prepared for their duties through their own efforts, and that they would demonstrate their proficiency by their performance on the tests. By the year 2200 B.C., key public officials were being examined once every three years. After three examinations, the official was either promoted or fired. There is no record of the content or methods used in these examinations, but the system seems to have worked very satisfactorily, because it was continued for many generations.

In 1115 B.C., at the beginning of the Chan dynasty, formal examinations were instituted for candidates for public office. The tests were "job samples" requiring that candidates demonstrate proficiency in the five "basic arts"—music, archery, horsemanship, writing and arithmetic. Knowledge of a sixth art was also added—skill in the rites and ceremonies of public and social life. Millions of people prepared for the tests, but relatively few achieved success. Over its long history, the procedures were changed from time to time, with a test on moral standards included about 165 B.C. At various times the tests also included geography, civil law, military matters, agriculture, and the administration of revenue.

By 1370 A.D. three levels of examinations were well-established: district ("Budding Genius"), provincial ("Promoted Scholar"), and national ("Ready for Office"). Those who passed at each level received suitable honors, but only the very few who passed the final examination received a position and a seat in the grand council on the Cabinet, thus becoming a "Mandarin." In the early 1900's it was recognized that China lagged behind the West in military matters. This was thought to be because government officials lacked knowledge about science and technology, since the qualifying tests did not cover these subjects. The attempted solution was to establish universities and technological institutes, but people aspiring to public office were not attracted to these institutions, preferring the old examination system with its emphasis on the arts and literature. So the examinations were abolished in 1905 as a reform measure. (Dubois, 1964, 1970).

Early Performance Tests

Contrary to much of the current rhetoric, performance testing is very ancient. One well-known performance test is reported in Judges 12:5-6 of the Bible:

> . . . And when any of the fugitives of Ephraim said, "let me go over," the men of Gilead said to him,

> "Are you an Ephraimite?"
>
> When he said, "No," they said to him, "Then say Shibboleth," and he said "Sibboleth," for he could not pronounce it right; then they seized him and slew him at the fords of the Jordan. (The Holy Bible, Revised Standard Version)

Another example of a performance test with very high stakes was the "floating test," used by colonists in Salem, Massachusetts, to detect witches. Suspected witches were dunked into water. Those who floated were declared to be witches and were burned at the stake; those who sank and drowned were declared not to be witches.

University Examinations in Europe

The first records of examinations in education appeared in the Middle Ages in European universities. The earliest examinations were oral law examinations at the University of Bologna in 1219. Louvain University had competitive examinations in the mid 1400's to place students in the following categories: honors, satisfactory, charity passes, and failures. The early examinations were oral, but by 1803, use of written examinations was widely accepted (Popham, 1981).

NINETEENTH CENTURY DEVELOPMENTS

The Chinese civil service examination system was much admired in Europe, and much of the impetus toward civil service testing there and in the United States was based initially on the Chinese system. Westerners were impressed with the fact that competition was open, that distinction came from merit, and that a highly literate and urbane group of public officials resulted from the examination system. The idea of an examination system for civil service employees had spread to Europe by the 1800's, and in the 1850's the British began their own system. By the 1860's, interest had spread to the United States and a formal Civil Service Board was established in 1871 by President U.S. Grant. The assessments used by the U.S. Civil Service Board included short answer items, biographical information, and a six-month on-the-job performance rating (Popham, 1981). Three years after the establishment of the Board, criticism was heard that the examination, while helpful in selection, did not really predict on-the-job performance (Dubois, 1970). This issue is still a problem for employment and licensure examinations.

Development of Experimental Psychology

Modern psychological and educational testing developed from several movements that occurred during the nineteenth century. The primary impetus for these developments was the application of the "scientific method" to the study of human beings. In psychology, the power of the biological and physiological sciences based on the scientific method caused psychologists to abandon their philosophical leanings and begin to look for "hard indicators" of psychological traits. The first experimental psychology laboratory was established in 1879 by Wilhelm Wundt at Leipzig. One of his students, James Mckeen Cattell, an American, returned to the United States in the late 1880's and began to spread Wundt's ideas. One of Cattell's students, E.L. Thorndike, became the "father of modern psychological and educational testing" through establishment of a department at Teachers College at Columbia University in which the early leaders in educational and psychological measurement were trained.

The psychological laboratories instituted strict controls on a variety of factors that could affect behaviors to be examined. Measurement of learning, for example, required objective indicators of change. Standardized items and administration became a necessity. Sir Francis Galton applied these ideas to the study of differences among individuals as to physical and psychological traits. In 1882, he established a laboratory in London, England, where he explored individual differences among related and unrelated individuals. Objective measures were required because comparisons among individuals were essential to the research. He also pioneered much work in the development of rating scales and questionnaires (Anastasi, 1961).

The work of Darwin in 1859 also contributed to the interest in objective measures, because his work focused on differences among members of a species and required comparable data for comparisons (Thorndike & Hagen, 1955).

Early Achievement Testing in Education

Prior to 1850, educational appraisal relied almost totally on oral examinations conducted without attention to standardization or uniformity in questioning or procedure. While such procedures allowed teachers opportunities to determine some of what individual students had learned, the results were very inconsistent and provided no information for comparison of students. These comparisons were desired for making such decisions as selection of students for entrance into academies and colleges (Thorndike & Hagen, 1955).

In order to develop some basis for comparing students, oral examinations were largely replaced by written essay examinations. These had advantages over the oral tests in that they allowed presentation of the same tasks to all students and each pupil could use a full examination period to formulate and record responses. However, the essays also presented problems. Readers of the essays often used different criteria for different students, changed their standards for model answers, or were influenced by extraneous factors such as handwriting or length of the responses.

An early attempt at standardization of scoring criteria was developed in the mid 1860's by Rev. George Fisher of Greenwich, England. Fisher collected samples or "specimens" of students' academic performance in the areas of writing, spelling, mathematics, and grammar and composition. These specimens were arranged in a "scale" book with values assigned to each specimen using a scale ranging from "1" (the best) to "5" (the worst). This was the first use of "anchor papers" for scoring (Cadenhead, K. & Robinson, R., 1987).

Joseph Mayer Rice, a pioneer in achievement testing, was interested in applying research methods developed by the psychological laboratories to the improvement of education. During the late 1880's, Rice was exploring methods to improve school efficiency. He started by personally administering a standardized spelling test to 33,000 children. He later developed and administered such tests in arithmetic, and language. His intent was to gather information about the effect on achievement of such variables as "amount of time." An important feature of Rice's work was that by administering the tests to large numbers of students over several grade levels, he was able to develop academic expectancies for each grade level, i.e., "grade equivalents." (Dubois, 1970).

All of the developments of the nineteenth century came together in the United States in the early part of the twentieth century and led to the development of modern measurement practices. James McKeen Cattell, a student of Wundt and a discipline of Galton, continued to explore individual differences in sensory and motor performance. In his article of 1890, he first used the term "mental test." These mental tests included assessment of reading, verbal association, memory, and arithmetic, although these tests often yielded conflicting results. The public's interest in such tests was piqued at the 1893 Columbia Exposition in Chicago. During the Exhibition, Joseph Jastrow set up an exhibit at which people could voluntarily test their own sensory, motor, and perceptual skills and compare themselves with "norms." Cattell's interest in studying individual differences was continued by his student, E.L. Thorndike who, with his students at Columbia University, fostered the standardized educational testing movement in the United States (Anastasi, 1961).

TWENTIETH CENTURY DEVELOPMENTS

Educational Research

In 1903 Rice established the Bureau of Research and began to publish *The Forum*, which reported scientific studies of education. The name of the Bureau was soon changed to the Society of Educational Research, which became the forerunner of the American Educational Research Association, organized in 1915 as the National Association of Directors of Educational Research. E.L. Thorndike was one of the contributors to *The Forum* (Dubois, 1971).

Intelligence Testing

In the early part of the twentieth century, humanitarian concerns for social and psychological "deviates" spawned an interest in describing the amount and type of deviation (Thorndike & Hagen, 1955).

Most of the early psychological studies involved only sensory and motor characteristics, although the investigators were looking for indicators of intelligence. It was the recognition that the simple functions being studied provided little useful information that led Alfred Binet and Theodore Simon to study the end products of intellectual functioning and so to develop the Binet scales. They devised a series of tasks for testing children who were not doing well in school in order to describe the nature of their difficulties.

The Binet-Simon instrument, published in 1905, contained 30 problems (items) arranged in order of difficulty. The tests included verbal, sensory, and perceptual tasks and yielded a "Mental Age" score. Two revisions published in 1908 and 1911 expanded the first test. The Binet-Simon test was brought to the United States and was published in 1916 by Lewis Teran of Stanford University as the Stanford Binet Intelligence Test. Originally, the test employed many of Binet's concepts and test items. In 1937 and again in 1960, the Stanford Binet was revised by Terman and Maud Merrill. Since then, there have been many other revisions, normative data have been collected, and statistical methods have been applied to investigate the qualities of the instrument (Thorndike & Hagen, 1955).

Group intelligence tests developed from work on the Army Alpha and Army Beta tests, which were used to classify men for branches of the U.S. military service in World War I. These tests were developed after the American Psychological Association set up a committee, headed by Robert Yerkes, to determine how the Association might help in the war effort. The tests were based on the kind of tasks used in the Stanford Binet, and they featured the use of multiple-choice items, which had been developed by Arthur Otis, a student of E.L. Thorndike. The Army Alpha and Beta provided information which was used to make personnel decisions such as rejection or discharge, assignment to service, and admission to officer training. The Army Alpha test was a general purpose test, and the Army Beta test, a nonverbal test, was developed for use with the foreign born and illiterates. The Otis Intelligence Test, derived from the Army Alpha, was later released for public use and Americans took the tests in large numbers. Scores were reported as an Intelligence Quotient (IQ), the ratio of the mental age to the chronological age. Thus, began the American fascination with "IQ." (Anastasi, 1961)

Four other individual intelligence scales were developed during the period from the 1930's through the 1960's by David Wechsler and his associates.

The Wechsler Bellevue Scale was developed in the 1930's and was widely used during the late 1940's for testing veterans of World War II. The Wechsler Intelligence Scale for Children (WISC) was published in 1949, the Wechsler Adult Intelligence Scale (WAIS), a revision of the Wechsler-Bellevue, was published in 1955 and the Wechsler Preschool and Primary Scale of Intelligence (WPPSI) was published in 1967. Each of these tests has a separate scale for verbal and performance tasks; in addition, each of the scales has numerous subscales which are separately scored to provide a diagnostic profile in addition to verbal, performance, and total IQ. (Nunnally, 1972)

Educational Achievement Tests

During the 1920's, many people used techniques employed in the development of the Army Alpha and Beta tests to develop group tests of intelligence and academic achievement in many subject areas. Test publishers developed and distributed "achievement test batteries," which covered such subjects as reading, arithmetic, language usage, social studies and science. Norms were developed to establish the "grade equivalent" of scores on the various tests.

Scoring and Analysis Procedures

The development of test items that could be scored completely objectively, that is , with no judgment on the part of the scorer, is attributed to Arthur Otis, whose multiple-choice items were used with the Army Alpha. Use of these items reduced the scoring time and made possible large-scale testing programs. During World War II, a test scoring machine was developed, further reducing the time needed to score large groups of tests. More recently, development of electronic data processing equipment has greatly expanded the possibilities for scoring, analysis, and preparation of tests. Other technical advances have made it possible to administer tests by computer and have them scored immediately.

PERIODS OF DEVELOPMENT AND CONCERN

Within the twentieth century, there have been four distinguishable periods of test development. The first period, from 1900 to 1915, involved pioneering work. The period from 1915 to 1930 saw the rapid development of techniques for group testing in intelligence and achievement, as well as other types of assessment devices. From 1930 through the 1960's there was a great expansion of the use of tests. This period also included the development psychometric theory, along with the beginning of criticism of the tests and of the uses made of them. At the present time, tests are widely used and are frequently seen as the vehicle for change in education; however, serious questions about the effects of the testing and the need for making changes in the procedures have been widely advocated.

A recap of each of these eras is presented below.

Pioneering Period: 1900-1915

At the turn of the century, faculty psychology guided much of the curriculum development effort and almost all of the test development. The mind was conceived of as having certain faculties, such as memory and reason, which were to be trained.

In France, Binet and Simon pioneered the testing of intellectual tasks, from which the early group tests of intelligence were derived. Also during this period, standardized educational achievement tests began to appear. These included Stone's 1908 arithmetic test, Thorndike's 1910 handwriting scale, Buckingham's 1915 spelling tests, and Trabue's 1916 language tests. In addition, Thorndike used Rice's technique of setting expectancies based on actual performance of students from different grade levels. He also adapted Fisher's ideas of "specimen" papers to set criteria for the score categories. (Dubois, 1970).

Rapid Development: 1915-1930

As an aftermath of the Army Alpha and Beta tests, the years from 1915 to 1930 saw a tremendous growth in the numbers of tests and achievement batteries made their appearance.

The growth of behavioral psychology in the 1920's impacted on the characteristics of the new tests. Behaviorists believed that people learned best when tasks were broken down into smaller skills that could be arranged in hierarchies. This idea, emphasizing basic facts and skills, spread to education and to teaching methodology and, consequently, to testing. This influence on testing lasted until the late 1980's (Stiggins, 1991).

In addition, tests of other variables began to appear. Rorschach introduced his projective personality test using inkblots in 1921. Seashore developed a music test in 1919 and Strong published the first vocational interest test in 1927 (Dubois, 1970).

Rapid Expansion; Beginning of Criticism: 1930-1960

Tests became an accepted part of the American culture in education and psychology and a relatively quiet period in American education continued through 1959 (Findley, 1963).

Use of tests received a great impetus during World War II. In 1943, the Office of Strategic Services, or OSS (later to become the CIA) developed and used personnel tests which measured such variables as planning and handling stress. These tests used scenarios and real-life role-playing situation to assess recruits' ability to plan and carry out military intelligence work. In addition, the old Army tests were expanded into the Army General Classification Test. It included reading vocabulary, mechanical, clerical, code learning, and oral trade. Other tests were developed to select persons for specific military jobs such as pilots, gunners, and navigators (Dubois, 1970).

This work led to the development of tests of specific aptitudes, with the expectations that such tests could assist in selection of careers and occupations. The General Aptitude Test Battery (GATB), developed by the U.S. Employment Service, has been widely used and some validation data has been collected. The Differential Aptitude Test (DAT) was developed in 1966 for use in counseling programs in the schools. In 1967, Guilford and his associates published a series of papers dealing with the Structure of the Intellect.

This period also saw a change in emphasis in testing. One change was the return to more global assessments of a large range of educational skills.

The field of psychology also changed its approaches with more global projective assessment methods. (Thorndike & Hagen, 1955). Findley (1963) also noted that this period saw the development of new basic skills and reasoning tests and the wide-spread use of "power" instead of "speeded," or timed, tests.

During this period, psychometric theory was developed and concepts of standards were developed. As part of the movement to foster more critical use of tests and improved test quality, Oscar Buros began publishing a new work in 1935 called the Mental Measurements Yearbook. This book contained a listing of tests in print and, beginning in 1937, included critical reviews of the listed instruments. The Yearbooks have been issued approximately every five years, with the eleventh Yearbook being published in 1992.

The first edition of Educational Measurement, edited by E.F. Lindquist, was published in 1951 and the second edition, edited by R.L. Thorndike, was published in 1971. These publications attempted to bring together in one volume what was known about all aspects of testing. The third edition, edited by R.L. Linn was published in 1989. Each of these volumes provides a comprehensive guide to educational measurement as it was practiced at the time of publication. All have been sponsored by the American Council on Education.

Basic psychometric concepts such as validity and reliability and other key psychometric issues were discussed in books and journals. In 1954, the American Psychological Association, the American Educational Research Association, and the National council on Measurement in Education published the first Standards for Educational and Psychological Tests. The standards were revised in 1966, and again in 1985 (APA, 1985). More recently, standards for training teachers in assessment have also been developed (Joint Committee on Testing Practices, 1988).

In 1959, an event that greatly influenced American education occurred. The USSR launched an earth satellite called Sputnik and launched American education into a period of self-doubt and self-criticism that has lasted to the present time. Many people questioned the quality of American education, asking such questions as, "How could the excellent education system that we thought we had fail to produce the first of such satellites?" The entire educational process was subjected to critical scrutiny and many educational "reforms" were inaugurated. The question for educational testing was: Why didn't our testing show us the "inferiority" of our schools? The solution to this dilemma was to advocate more testing.

During this period, enthusiasm for assessment was so strong that misuses became common. Test results were often accepted as totally definitive indicators of achievement and intelligence. Scores were used for many inappropriate purposes and criticism began to mount. Heredity-environment issues were hotly debated, students were often grouped and tracked solely on the basis of test scores. Questions were raised bout the limited scope of content and skills covered on the tests, and even as to the efficacy of the underlying philosophy on which the quantification of behavior was based. Such criticisms are still with us today.

Continued Expansion: Mounting Criticism: 1960 to Present

This has been a period of many contradictory movements, many of them political in nature. The greatest contradiction has involved continuing requirements for testing on the part of the public and governmental agencies, while criticisms of tests have grown increasingly vitriolic. For a time broad coverage and norm-referenced interpretations were viewed as the source of the problems, so narrowly specified objectives and objective-referenced interpretations were viewed as the solution.

Recently, however, there is advocacy for broader, more global objectives and for use of more varied assessments. This period began in the early 1960's when the Elementary and Secondary Education Act (ESEA) was passed into law. The act required that all schools receiving federal aid to education show evidence that they were accomplishing educational goals with these monies. The federal government specified the kinds of evidence to be accepted as proof of the effectiveness of federally funded programs. The evidence was to consist of results of educational tests administered to recipient children in schools. School districts were required to administer tests, analyze the results, and report them to Washington.

The interest in educational change programs such as ESEA widened to include programs for limited English speakers, minority persons, the gifted, the physically and mentally handicapped, and preschool children and adults. All of these programs required program evaluations that centered on test data. Thus, tests had to be administered often and for a number of purposes. Teachers became the primary administrators of such tests, even though most of them had little or no training in test use and test construction.

Along with requirements for more testing for program evaluation, there was also an interest in determining the "picture" of educational progress within the U.S. on a national basis. In 1969, Federal legislation was passed to begin the National Assessment of Educational Progress (NAEP). NAEP tested a national sample of students aged 9, 13, and 17 and adults aged 25 to 36 in the areas of reading, mathematics, writing, science, citizenship, literature, social studies, career and occupational development, art, music, history, geography, and computer competence. Items included short answers, essays, observations, questions, interviews, performance tasks, and sample products (Gronlund and Linn, 1990).

In the 1970's there was a shift of focus from the federal to the state level and the advance of the accountability movement and minimal competency testing in education. During this period, an increasingly disgruntled public had become impatient with what were p perceived as small educational gains and demanded clear "evidence" of educational attainments. The evidence was to be in the form of tests and test scores, and high stakes were attached to the results (Jaeger & Tittle, 1980).

Schools, districts, and states began the development of their own tests or used commercially developed instruments to make schools "accountable" for their use of public funds to educate children. The results were frequently published in the press with the names of schools and districts prominently displayed. However, in many states, teachers were seldom able to use the test results for the purposes of improving their instruction or diagnosing children's needs.

The most widely used educational tests are standardized and norm-referenced, that is, students are tested on a common broad knowledge base and their scores are compared or referenced to those of a wide variety of students from many locales. Part of educators' response to the public's demands for accountability has been to criticize the tests used. Beginning in the early 1970's, the use of tests has continued to grow while the criticisms have become even more insistent.

One criticism has been that nationally standardized tests are deliberately designed to be so general in content coverage that they do not reflect the curricula that teachers in any specific school district are teaching. Another criticism came from behavioral learning theorists who saw the standardized, norm-referenced, tests as being incomplete and inconsistent in examining skills. Some of this disenchantment was a consequence of the use of programmed instruction and teaching machines, which break learning into small, sequential steps in order to "guarantee" students' mastery of the concepts under study. Students taught by these methods sometimes failed to show mastery on the standardized, norm-referenced tests (Glaser, 1963).

One solution offered for this problem was to develop a new form of test, one that measured exactly what was to be taught, rather than more general goals. These tests were dubbed "criterion-referenced tests" by Glaser in 1963. They examined students on very carefully specified content. The content often came from lists of skills, such as those for "minimal competence" (Popham, 1978). Scores were interpreted in terms of mastery of objectives, rather than normatively. That is, students were compared, not with other students across the country, but with some standard of mastery of the specified content. The criteria were in the form of cut scores of total amounts of content mastered or of mastery of individual skills. If a student's score was above the cut score he or she was considered "competent." If the student's score was below the cut score, the student generally received remedial instruction on the "missed" item content.

Test results for minimally competent students were publicly announced, and many school districts began to spend a large part of their resources and classroom time working on the "minimal" knowledge and skills to the detriment of the rest of the curriculum. Furthermore, because the measured skills and knowledge were often tested and taught as discrete pieces of information, students were unable to put the knowledge and skills together into cohesive, useful bodies of learning. Schools were producing readers who could identify individual words but could not comprehend connected text, or who could recognize good grammar and syntax. but could not write coherent paragraphs. Experts from the fields of mathematics, science, and the social sciences reported similar problems. The old criticism of narrow content had re-emerged.

In the 1980's, criticism focused on the almost exclusive use of multiple choice tests. Critics insist that a different type of assessment should be used. Ironically, many of the "alternative" assessments being advocated are very much like those of earlier history, with the same problems that were found with them, although many of the advocates have seemingly been unaware of this.

The assessments being currently recommended are termed "authentic" or "alternative." This emphasis began during the 1980's, largely at the insistence of cognitive psychologists. These psychologists feel that children learn best with content and procedures grouped and organized (Lane, 1989; Glaser, Lesgold & Lajoie, 1987). They also feel that assessments that fragment learning do not accurately assess what children really know.

The cognitive point of view is reflected in such changes as the return to the teaching of writing as a process, asking students to actually write, rather than to simply recognize good writing or components of it. Mathematicians have also begun to look at the teaching of mathematics from the "constructivist" perspective. That is, students have to construct their own knowledge of how math works.

Teachers have been urged to change their modes of teaching, de-emphasizing the lecture and increasing the use of projects, experiments, and productive tasks. However, some critics feel that, when students taught by the "new" methods are tested by traditional tests, the "new" types of student learnings are not always reflected in higher test scores.

Tests are also criticized as to the types of tasks and item formats used. Critics have stated that the tasks, typically multiple choice, assess only surface learning of facts. Shavelson, Baxter, & Pine (1992) called such test items "surrogates" to the real tasks of students. Some educators have begun to demand tasks that are "alternatives" to the more traditional multiple-choice tests, tasks that are more similar to actual or "authentic" work of students. Such tasks involve samples of writing, completion of experiments, production of products such as reports and speeches, and problem-solving activities.

Today, some educators and educational agencies are busily developing "authentic" or "alternative" assessments (Wiggins, 1975). The new tasks are more appealing to students and teachers, but there are problems. Many of the items are poorly written and, in spite of the intentions of the writers, tap only low-level skills. In addition, the same problems that led to the eager adoption of multiple-choice items in the 1920's have resurfaced; i.e., bias in scoring, unreliability of scores, poor model answers, unclear scoring criteria, and introduction of irrelevant factors such as response length and handwriting.

POSTSCRIPT

Can we use lessons from the past to help us develop and use truly authentic assessments—Assessments that provide unambiguously and dependably the kind of information needed for the purpose for which they are used?

As a minimum, assessments should—

fit the purpose for which they are used.
cover important content.
require that the intended skills be demonstrated.
present the same tasks to every examinee.*
use the same scoring criteria for every examinee.
yield consistent results.
be fair to every examinee.

We know how to do this for multiple-choice tests—although too often we fail to do the things we know how to do. For other types of assessments, we often do not even know how. There is much work to be done.

*Except in the case of "Matrix Sampling," which is used for some survey testing.

Table 1. Events and Dates in the Development of Educational Testing

MAJOR CONCEPTS	DATES
China	2200 BC - 1905 AD
European Civil Service	1850
U.S. Civil Service	1871
Darwin	1857
Psychological Lab	1879
Galton	1882, 1890
Cattell	1880
Fisher's Scale Book	1865
Rice's Tests	1887
Stone's Arithmetic Test	1908
Thorndike's Handwriting Test	1910
Buckingham's Spelling Test	1915
Trabue's Language Test	1916
Achievement Test Batteries	1926

Bureau of Research (Rice)	1903
AERA	1915
Binet and Simon	1905, 1908, 1911
Terman and Merrill	1916, 1937, 1960
Army Alpha and Beta	1917
Otis	1921
Wechsler-Bellevue	1939
WISC	1949
WAIS	1955
WPPSI	1967
Army General Classification Test	1943
GATB	1946
DAT	1966
Rorschach	1921
Strong Vocational Inventory	1927
Objective scoring	1915
Buros MM Yearbook	1935 (1st), 1992 (11th)
Beginning of development of psychometric theory	1940's
Publication of Educational Measurement	1951, 1971, 1989
APA Standards	1954, 1966, 1985
Code of Fair Testing Standards	1988
Faculty psychology	1850 - 1920
Behaviorism	1920+
Generalization	1963+
Cognitive psychology	1980+
Sputnik	1959
ESEA	1960
Minimal competency programs	1960
Authentic Assessment	1989
Alternative Assessment	1989

REFERENCES

American Psychological Association, American Educational Research Association, & National Council on Measurement in Education. (1985). Standards for educational and psychological testing. Washington DC: American Psychological Association.

Anastasi, A. (1961). Psychological testing. New York: The Macmillan Company.

Cadenhead, K. & Robinson, R. (1987). Fisher's "Scale-book": An early attempt at educational measurement. Educational Measurement: Issues and Practice, 6(4), 15-18.

Dubois, P.H. (1964). A test-dominated society: China, 1115 B.C. - 1905 A.D. Proceedings of the 1964 Invitational Conference on Testing Problems. Princeton: Educational Testing Service.

Dubois, P.H. (1970). A history of psychological testing, Boston: Allyn and Bacon, Inc.

Dubois, P.H. (1971). Increase in educational opportunity through measurement. Proceedings of the 1971 Invitational Conference on Testing Problems. Princeton: Educational Testing Service.

Findley, W.G. (1963). Purposes of school testing programs and their efficient development. In W.G. Findley, (Ed.) The impact and improvement of school testing programs: The sixty-second yearbook of the National Society for the Study of Education, Part II, Chicago: The University Press of Chicago Press.

Glaser, R. (1963). Instructional technology and the measurement of learning outcomes—some questions. American Psychologist, 18, 519-21.

Glaser, R., Lesgold, A. & Lajoie, S. (1987). Toward a cognitive theory for the measurement of achievement. In R.R. Ronning, J.A. Glover, J.C. Conoley & J.C. Witt (Eds.). The influence of cognitive psychology on testing. Hillsdale, NJ: Erlbaum.

Gronlund, N.E. & Linn, R.L. (1990). Measurement and evaluation in teaching. New York: Macmillan Publishing Company.

Jaeger, R.M. & Tittle, C.K. (Eds.) (1980). Minimum competency achievement testing: Motives, models, measure, and consequences. Berkeley, California: McCutchan Publishing Company.

Joint Committee on Testing Practices. 1988. Code of Fair Testing Practices in Education. Washington, DC: American Psychological Association.

Lane, S. (1989). Implications of cognitive psychology on measurement and testing. Educational Measurement: Issues and Practice, 8(1), 17 - 19.

Nunnally, J.C. (1972). Educational measurement and evaluation. New York: McGraw-Hill

Popham, J. (1978). Criterion-referenced measurement. Englewood Cliffs: NJ: Prentice-Hall, Inc.

Popham, J. (1981). Modern educational measurement, "Englewood Cliffs, NJ: Prentice-Hall, Inc.

Shavelson, R.J. , Baxter, G.P., & Pine, J. (1992). Performance assessments: Political rhetoric and measurement reality. Educational Researcher, 21(5), 23 - 27

Stiggins, R.J. (1991). Facing the challenges of a new era of educational assessments. Applied Measurement in Education, 4(4), 263 - 273

Thorndike, R.L. & Hagen, E. (1955). Measurement and evaluation in psychology and education. New York: John Wiley & Sons, Inc.

Wiggins, G. (19750. Discussion, Educational indicators: Monitoring the state of education. Proceedings of the 1975 Invitational Testing Conference. Princeton: Educational Testing Service.

Part 2

The Portfolio: Scrapbook or Assessment Tool?

A paper presented at the William F. Breivogel Conference,
University of Florida, March 25, 1993

By: Dr. Jonnie P. Ellis, Supervisor Chapter I
Alachua County

Mrs. Jeannette Hiebsch, Chapter 1 Teacher
Stephen Foster Elementary School

Ms. Shirley Taylor, Chapter 1 Teacher
Lake Forest Elementary School

Abstract:

A case study of how a Kindergarten-First Grade Chapter 1 program changed a collection of classroom worksheets into a sound assessment tool for measuring academic growth in an instructional setting. Portfolio purpose, design, content, management, data synthesis-analysis, and storage issues were addressed by Chapter 1 personnel in making this transition.

In the 1991-92 school year kindergarten-first grade teachers began keeping portfolios for their students. Two factors in the educational arena led to this decision by Chapter 1 teachers. First, reliance on achievement test scores as the data source for program evaluation and as the primary measure of student growth had been a common concern for several years. Early childhood experts, over the years, have urged against over-reliance on achievement test scores for five and six year old students as a basis for determining academic progress. In fact, they have continued to question the appropriateness of even using achievement tests for children. A second factor, the 1989 reauthorization of the Chapter 1 legislation, increased the level of teacher concern. That legislation focused attention on a new issue - "Program Improvement." This issue brought more stringent criteria for measuring academic success both at the program level and for individual students. As a result, numerous schools in our district didn't meet the DOE suggested criterion and were, therefore, flagged for program improvement. Those schools were required to write a plan to improve the Chapter 1 program for kindergarten and first grade students in reading and/or math. A second year program flag required implementation of the "approved" plan and a third year flag brought technical assistance from DOE personnel. At this point, DOE personnel must assist in the development of the required plan and provide increased monitoring to ensure that the plan is appropriately implemented.

In addition to the program improvement planning process, Chapter 1 legislation required that students not reaching the "suggested" criterion be identified so that they can be targeted for additional instructional support. This resulted in students going on to second grade labeled as "Chapter 1 first grade failures." The dilemma teachers faced was in proving to themselves and second grade teachers that their students had achieved academic growth and experienced success in first grade, even though they had not met the criterion for success on the achievement test at the end of the year.

While this problem was being discussed at Chapter 1 workshops another event occurred. This state schoolwide accountability and program improvement legislation required each local school to take a closer look at achievement of all students. Local school teacher-parent committees began making plans to modify, and, in some instances, to restructure the schoolwide instructional environment to meet goals and standards outlined in Blueprint 2000. The time line for implementation of new assessment procedures as outlined in that document suggested a possible solution to the Chapter 1 first grade dilemma. The "Assessment Portfolio" seemed to be the answer. The timing was perfect! The portfolio would provide a place to store performance samples to document academic growth of students. Chapter 1 teachers agreed to get a headstart on Blueprint 2000 portfolio assessment because simply storing the work samples they were already collecting would not require additional time. And they were sure when someone "thumbed" through the portfolio collection the academic growth of students would be evident.

In a training session, teachers discussed what a student portfolio should contain. Information and work samples showing the intellectual, physical, emotional, and social development were listed as minimum expectations for what should be included in a portfolio collection. The most frequently used checklists were identified and instructions for their use were provided for teachers to review. It was suggested that teachers choose several checklists and at least two cut-and-paste activities to include in a student's portfolio. It could also include any other items the teachers chose.

After about three months, the teachers became very critical of the portfolios they received when students transferred from one school to another. Each teacher measured a "transfer" portfolio against her expectations and of course each teacher's expectations became the set of expectations with integrity. Teacher comments were varied. "The only thing in the portfolio was arts and crafts!" "Initial consonant sounds were checked as mastered for a student who doesn't even recognize alphabet letters." "Not a single math checklist or work sample was included!" Negativity about the portfolios continued to increase as teachers shared their opinions and comments at the regularly scheduled inservice sessions. The final blow came from a teacher who said, ". . . the portfolio I received was nothing but a scrapbook of cutesy artsy cut-and-paste worksheets." It was time to take action. A committee was appointed to study the issue and report their findings to the group. It was suggested that the committee review the current research related to portfolio assessment and come back to the group with recommendations to improve the portfolio or drop it altogether.

Portfolios: What and Why?

The task had been defined. Review the research and answer the questions posed. What is a portfolio? Why do one? What items should be included in it? How can its contents be used to assist teachers with instruction and document student growth?

A wide range of descriptions were discussed as the committee attempted to reach consensus in defining the term "portfolio." Flood and Lapp (1989) described a portfolio designed to report reading progress to parents as "... a comprehensive comparison report that includes grades, norm-referenced tests, criterion-referenced tests, and informal measures." Some members of the committee were surprised that reading experts supported the inclusion of norm-referenced test data in students' portfolios. An article by Olson (1991) helped the committee to distinguish between what Chapter 1 teachers were doing and what they actually wanted to do. Olson said a folio is ". . . .a large collection of materials relating to one or more dimensions of a person's educational or professional life," and a portfolio is ". . . .a smaller collection of materials selected for a specific purpose." Chapter 1 teachers were maintaining student "folios" rather than portfolios! Clearly items included in the portfolio would have to be "selected" according to some - as yet undefined - criteria. "A purposeful collection of student work and records of progress and achievement collected over a period of time . . . .a tool for expanding the quantity and quality of information we use to examine learning and growth" was the way Valencia (1991) defined a portfolio. This definition focused attention not only on the quantity of information but, more importantly, the quality of the information to be included. It also suggested that its purpose was to examine learning and growth over time. Several other characteristic of portfolios are discussed by Arter and Spandel (1991).

> A (student) portfolio is a purposeful collection of student work that exhibits to the student (and/or others) the student's efforts, progress or achievement in (a) given area(s). This collection must include:
>
> (a) student participation in selection of portfolio content;
> (b) the criteria for selection;
> (c) the criteria for judging merit; and
> (d) evidence of student-reflection.

The "purposeful" collection of portfolio items gave further support to the development of criteria for selection. Establishing criteria for judging quality of the performance samples identified another dimension of the portfolio assessment task for the committee to address.

Setting standards for judging the quality of the performance (Arter and Spandel, Spring 1992) is critical if a collection of performance samples is to have assessment value. Proponents of portfolio assessment emphasize the value of setting standards because it addresses growth over time. Teachers have a professional obligation to evaluate student performance against specifically defined criteria, that can be clearly articulated. However, it is difficult to specify benchmark activities and appropriate performance samples to quantity success or developmental stages that fit easily into a portfolio in some areas. Individual records based on teacher observations and interpretations of student behaviors must be based on predetermined criteria to reduce the influence of "subjectivity" on the data included. The committee found little research to clarify this issue. The teachers were particularly concerned with the "expertise" of the collector, since the data in Chapter 1 folders could have been collected by the teacher or classroom aide. While it may be appropriate to include some samples collected by the aide, the feeling was that some items should be designated for the teacher only.

Even though the committee found little research related to the quality of the response mode, this is a very important variable in determining the validity, reliability, and fairness of portfolio items. It cannot be assumed that performance samples are bias free. Nor can it be assumed that selected samples are aligned with and supportive of the existing subject area curriculum. Standardized procedures must be followed in collecting portfolio samples that provide valid and reliable measures of academic growth in the various subject areas. Time lines must be established and data collection points specified to document learning and growth over time. Verifying the quality of the response mode and studying comparisons of selected items to determine their validity and reliability will take time.

Another topic of concern in maintaining portfolios is that of organization. A collection of assessment items and performance samples stored in a haphazard manner cannot be used efficiently nor effectively to document student growth or appropriate instructional planning by the teacher. The members of the committee spent many hours in dialogue about how the Chapter 1 portfolio should or could be organized. Selections of performance samples to reflect achievement of the teacher's instructional goals was a factor in deciding how the portfolio should be organized. It needed to be organized in a way that facilitated the aggregation and analyses of raw data—a necessity in documenting program success. Summaries of evaluation data derived from such analyses are useless, though, unless data sources are valid, reliable and relatively free from subjectivity. Analyses that cause questions to arise must be traced to the raw data sources for clarification. When data is organized in some logical fashion difficulties of this sort can be more easily resolved.

To better understand the advantages of using portfolios in the assessment process, the committee reviewed articles describing the implementation process and the progress of portfolio evaluation at several sites involving public schools at the district and state levels. In a 1989 article, Flood and Lapp gave specific suggestions for documenting reading progress in a comparison portfolio that was designed for parents. Rather than rely on a single test score the comparison report included grades, norm-referenced and criterion-referenced tests, as well as informal measures. (See Figure 1)

Data included in this portfolio provided an opportunity to compare a child's performance from time to time. In other words, a child's academic growth was measured over time and shown in a way that was easy for parents to understand. (See Figure 2)

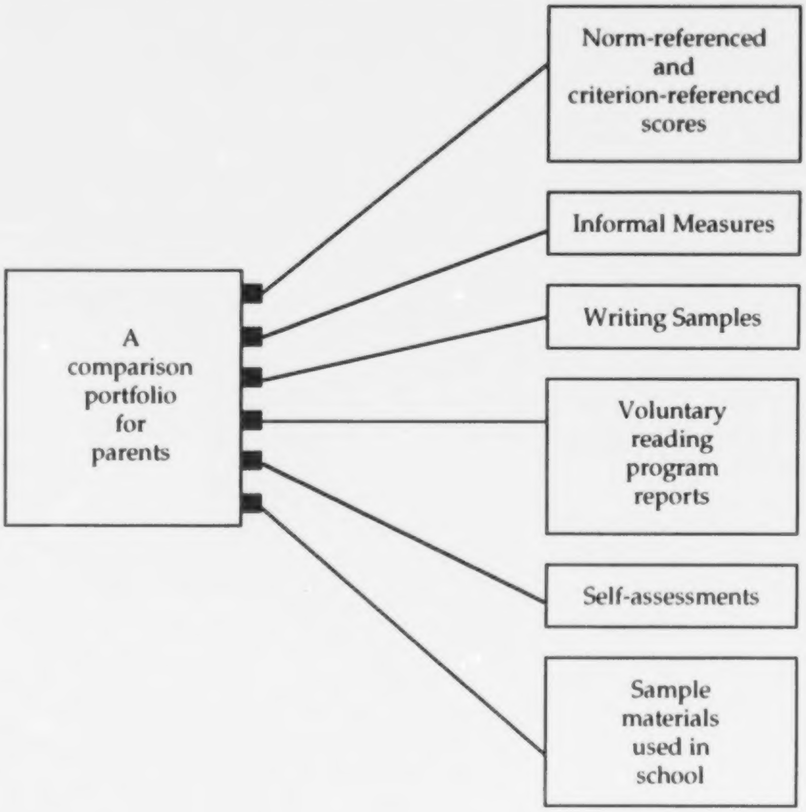
The advantages posed by Flood and Lapp were extended and elaborated upon at implementation sites reviewed by Olson (1991). The Department of Education in Vermont adopted the portfolio as an alternative assessment system to assess math and writing skills of public school students in that state. The Vermont project focused on the instructional value in portfolio assessment in three ways; by documenting student growth in writing over the span of a whole year, by giving a complete picture of a student's individual achievements, and by using it in parent conferences to better facilitate the development of math concepts through representations in charts and graphs and develop an understanding of the various steps in the writing process.

In the Kamehameka Elementary Program (KEEP) in Hawaii, development and growth of literacy skills is being monitored with portfolios. And in the Orange County Public Schools, Orlando, Florida, portfolios are being used to assess reading success of primary students. In another discussion of the Orange County portfolio implementation (Mathews, 1990), the crucial need for staff development, ongoing support, time lines, and standardized collection procedures are further described. In all three instances, assessment procedures have been developed based on the belief that assessment and instruction are linked—they support each other. Assessment provides teachers with the critical information necessary for instructional planning and curricular design. "What students know" was an important consideration in choosing items to include in the portfolio at each of the sites mentioned above. Portfolio assessment procedures must be grounded in the district's or, as in the case of Chapter 1, the program's need for accountability and the belief that the primary purpose of assessment is to provide teachers, students, and parents with information about academic growth of students.

The Matthews (1990) article stressed the importance of providing initial inservice training to prepare teachers to set up portfolio assessment procedures. This writer further emphasizes the critical need for ongoing support throughout the implementation process. Interestingly enough, the Vermont portfolio implementation has encountered serious problems (Rothman, 192) with low reliability resulting from inadequate teacher training. One teacher commented, "... There wasn't enough (teacher training) and it came too late." Apparently, initial training was minimal, not provided at the appropriate time and there was no provision for ongoing support. Realizing that the Chapter 1 portfolio assessment implementation began before teachers were adequately trained, mirroring the Vermont weakness, the committee turned its attention to the development of a plan to address staff development.

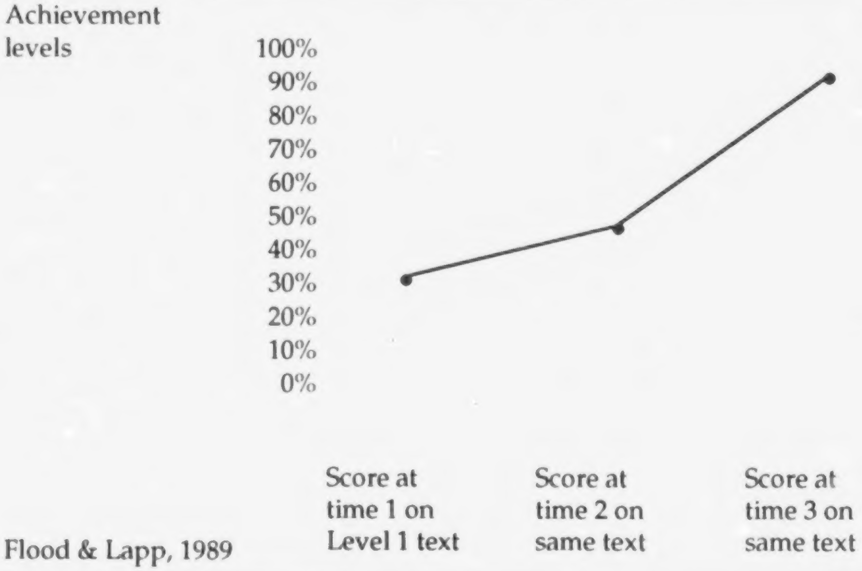
To transform a scrapbook of cutesy art projects into a valuable assessment tool demanded that teachers become expert observers of student behavior. This is an area often neglected in teacher education programs. Teachers must understand intellectual development of students. They must learn to use metacognition strategies such as the "think aloud" to help students become conscious, active, and more effective learners. Their knowledge about how the brain functions in processing information from targeted subject areas to revise existing cognitive knowledge structures or build more complex structures must be increased. Finally, the committee concluded that teachers can become expert evaluators only when they can integrate knowledge from all these areas to make informed instructional decisions. It was decided that the committee would plan a workshop session for teachers to share the research information they had collected. Once teachers were informed about the important issues and the procedures necessary to transform the portfolio scrapbooks into assessment tools, the committee planned to recommend that the portfolios be maintained. Then regularly scheduled sessions would focus on related issues and concerns throughout the implementation process. Members of the committee agreed that once initial training occurred, that setting up and maintaining portfolios, if properly supported, would be an effective vehicle for extending the teacher's expertise in all the critical knowledge areas identified areas (Olson, 1991).

Figure 1


A comparison portfolio for parents
Norm-referenced and criterion-referenced scores
Informal Measures
Writing Samples
Voluntary reading program reports
Self-assessments
Sample materials used in school

Flood & Lapp, 1989

Figure 2
Display of a child's growth in reading a single text



Flood & Lapp, 1989

Committee Action Plan

Since the committee had been appointed to study teacher concerns and make suggestions for improving or dropping the portfolio, they decided to set up new procedures for portfolio assessment. The new procedures were based on the committee's research in response to the questions raised and, therefore, would be the action plan record for the teachers to consider. The committee agreed on a purpose for setting up student portfolio assessment, suggested items to require and those to be considered optional. They set up an organizational structure, agreed on a time line for item collection, and listed specific conditions for collecting and recording data. The materials produced are included in Appendix A. They were presented to first grade Chapter 1 teachers as recommendations from the committee in an inservice workshop in August, 1992. The training session began with teachers working through a module called "Preparing for Portfolio Assessment" (Arter and Spandel, 1991). Various committee members shared the research information and described how the materials had been developed as teachers worked through the training module (see Appendix B).

During the current school year (1992-93) teachers have provided much feedback about the new procedures. They have suggested several procedural changes. They have discovered some mistakes - both typographical and content—that need to be corrected. A new committee will meet this summer to review the feedback, revise the procedures, and correct mistakes. They will also identify procedures to collect oral reading samples at several points during the year (Flood and Lapp, 1989). Interestingly enough, teachers are asking for more standardization—one of the things they objected to most about achievement tests. They demand strict adherence to the time line for collecting the data samples. They have discovered that Chapter 1 students who characteristically score low on the achievement test are achieving at lower than average levels on the performance samples included in their portfolios. However, the portfolio does document that over time the rate of growth for this group often exceeds that of average students. Because Chapter 1 students start out so far behind their peers an accelerated rate of growth must be established and maintained if they are ever to catch up!

The big disadvantage of portfolio assessments is the added burden on teacher time. Chapter 1 first grade teachers decided to continue portfolios and "find the time" required. Another potential disadvantage is space. Teachers are already predicting possible storage problems in the future. Finally, teachers are concerned that the Blueprint 2000 portfolio requirement will not be adequately funded. Money must be available to provide the staff development training and ongoing support necessary to make the portfolio a "valuable assessment tool." It must not become a scrapbook of cutesy art activities. Chapter 1 first grade teachers understand the time and effort necessary to transform a portfolio from a scrapbook into an assessment tool that can be used to effectively measure academic growth.

Appendix A

PORTFOLIO ASSESSMENT

Purpose/Rationale

The Chapter 1 K-1 portfolio will:

- Document student growth and development.
- Communicate with students, Chapter 1 teacher, LEA teachers, and parents.
- Guide instructional planning through the use of Chapter 1 checklist.
- Maintain standard assessment throughout the Chapter 1 program.
- Provide alternative program evaluation and program improvement data.

CHAPTER 1 K-1 PROGRAM
ALACHUA COUNTY

PORTFOLIO CONTENTS AND TIME FRAME

The following materials are required:

- Chapter 1 Assessment Record, K-1
  Kindergarten: October - optional, January, and the first two weeks of May
  First Grade: last two weeks of September, January, and the first two weeks of May.
- Chapter Checklist/Teacher Communication Card, K-1 (Four times a year)
- Prompted Writing Activity for **First Grade**
  (Last two weeks of October; first two weeks of May)

or

- Emergent Reading Assessment for **Kindergarten**
  (Last two weeks of October; first two weeks of May)

The following materials are optional:

- Reading Log
- Writing Log (Developmental Samples)
- Goal Sheet
- Journal
- Teacher notes
- Learning Log
- Teacher-student Conference Log
- Screening data - PREP, Speech/Language, SB/Ginn, MacMillan
- Cut & Paste & Color

KINDERGARTEN CHAPTER 1
EMERGENT READING ASSESSMENT

Name: ______________________________

(Use a different color pen to indicate second testing.)

Score: Total yes: _____/_____ Total no: _____/_____

Date: October _____/May _____

Comments:

Use with Who's Coming for a Ride? (Rigby, Stage 1, Set A)

1. Place book upside down on table. "Pick up this book and tell me what you think it will be about."

Yes ☐ No ☐ Finds front cover.

Yes ☐ No ☐ Able to make a sensible prediction.

2. "Where is the title?"

Yes ☐ No ☐ Finds the title.

3. "Read this book to me." If child says, "I can't read," use appropriate encouragement, such as, "It doesn't have to be like grown-up reading—just do it your own way." (If child refuses completely, go on to number 4.)

Yes ☐ No ☐ Uses pictures to tell story in own words.

Yes ☐ No ☐ Uses pictures to help with words.

Yes ☐ No ☐ Uses own language patterns.

Yes ☐ No ☐ Uses story pattern knowledge.

Yes ☐ No ☐ Uses beginning letter sounds.

Yes ☐ No ☐ Uses many letter sounds.

Yes ☐ No ☐ Accurately reads some of the words.

Yes ☐ No ☐ Self-corrects.

Yes ☐ No ☐ Recognizes some sight words.

4. When child is finished reading turn to pages 2-3. "I'm going to read this page to you. Point to the words as I read them."

Yes ☐ No ☐ Makes left to right motion.

5. Point to period, question mark, quotation mark. Ask student, "What's this?"

Yes ☐ No ☐ Identifies

Yes ☐ No ☐ Identifies ?

Yes ☐ No ☐ Identifies " "

6. "I want to keep reading. Where do I go next?"

Yes ☐ No ☐ Turns page appropriately.

7. Using page 5. Ask child, "Where does it say **pig**?"

Yes ☐ No ☐ Names the three letters.

8. "What letters are in the word **pig**?"

Yes ☐ No ☐ Names the three letters.

9. On the back of the page: "Write the word **dog**" "Write the word **cat**"

	dog	**cat**	
Yes ☐ No ☐	Initial Sound	Yes ☐ No ☐	Initial Sound
Yes ☐ No ☐	Ending Sound	Yes ☐ No ☐	Ending Sound
Yes ☐ No ☐	Entire word	Yes ☐ No ☐	Entire word

10. Point to exclamation point on page 8. Ask student, "What's this?"

Yes ☐ No ☐ Identifies

CHAPTER 1

PROMPTED WRITING ACTIVITY

Name: ____________________ School: ____________________

DIRECTIONS: Find the one description that fits the student's paper. The number next to the description is the score. If the paper is off topic, score the paper and write "Not Related to Prompt" at the top of the student's paper. Staple student's paper of this form.

Date: Comments/Evaluation

_____ 12. Student draws a picture related to a topic.
Student writes using some appropriate end punctuation and capitalization.
Student's writing exhibits structure appropriate to topic.
Student writes using some conventional spelling.
Student's writing conveys a sense of audience.

_____ 11. Student draws a picture related to a topic.
Student writes using some appropriate end punctuation and capitalization.
Student writes related sentences in a logical sequence.
Student writes utilizing correct and invented spelling.

_____ 10. Student draws a picture related to a topic. Student writes utilizing correct and/or invented spelling.
Student attempts multiple sentences to communicate ideas.
Students write using some appropriate end punctuation and capitalization.

_____ 9. Student draws a picture related to a topic.
Student writes using correct and invented spelling.
Student writes a sentence to communicate an idea.

_____ 8. Student draws a picture related to a topic.
Student writes using invented spelling.
Student attempts to write a sentence to communicate an idea.
In retelling, student exhibits an awareness of story.

_____ 7. Student draws a picture related to a topic.
Student writes using multiple consonants and some vowels.
Student uses invented spelling.
In retelling, student exhibits an awareness of story.

_____ 6. Student draws a picture related to a topic.
Student writes using initial consonants.

Date: Comments/Evaluation

In retelling, student exhibits an awareness of story.

_____ 5. Student draws a picture related to a topic.
Student writes using random letters.
Copies unrelated environmental point or uncorrected text.
In retelling, student exhibits an awareness of story.

_____ 4. Student draws a detailed picture that relates to a given topic.
Student attempts to produce writing in the form of squiggles.
In retelling, student exhibits an awareness of story.

_____ 3. Student draws a detailed picture that relates to a given topic or his/her retelling exhibits a sense of story.

_____ 2. Student draws simple picture relating to a given topic or relates a retelling.

_____ 1. Present during assessment but no attempt made in response to prompt.

Chapter 1 Assessment

Alachua County Chapter 1
Ed Smith, Director
Dr. Jonnie Ellis, Supervisor

Chapter 1 Assessment Record Instructions

A. The Assessment Record is to be used in conjunction with the Chapter 1 K-1 assessment binder.

B. Specific directions for administration of the assessment are included in the binder. We recommend a 3-5 second response time per item.

C. The Chapter 1 teacher fills in the following items on the record sheet:

1. name
2. LEA teacher
3. check (✓) either **K** or **1**
4. write in the year
5. circle items the child knows or answers correctly where appropriate
6. check (✓) areas of success for sorting and coins
7. complete math areas by recording the last correct number the child is able to count to or sequence

D. Note that kindergarten students will be assessed in January and May only. The September assessment column is to be filled out for 1st grade students in addition to January and May.

E. The "comments" column is a place for the Chapter 1 teacher to record anecdotal notes of interest. If you see this column, be sure to date your notes.

F. Attached samples are furnished on paper provided by the teacher. These can be stapled to the record sheet. We suggest one sheet of paper divided into thirds as follows:

First Name ______________________________

Last Name ______________________________

Numerals

Teacher writes date of sample just under student writing.

Chapter 1 Assessment Record

K ☐
1 ☐

Name: ______________________________ LEA Teacher: ______________________

READING

date:	September 199___	January 199___	May 199___
Recognizes capital letters out of order	knows: A B C D E F G H I J K L M N O P Q R S T U V W X Y Z	knows: A B C D E F G H I J K L M N O P Q R S T U V W X Y Z	knows: A B C D E F G H I J K L M N O P Q R S T U V W X Y Z
Recognizes lowercase letters out of order	knows: a b c d e f g h i j k l m n o p q r s t u v w x y z	knows: a b c d e f g h i j k l m n o p q r s t u v w x y z	knows: a b c d e f g h i j k l m n o p q r s t u v w x y z
Letter-sound correspondence (show picture, child says name, beginning sound, & letter name)	knows: A B C D E F G H I J K L M N O P Q R S T U V W X Y Z	knows: A B C D E F G H I J K L M N O P Q R S T U V W X Y Z	knows: A B C D E F G H I J K L M N O P Q R S T U V W X Y Z
Writes first name from memory	see attached	see attached	see attached
Writes last name from memory	see attached	see attached	see attached
Knows names of 8 basic colors	red orange yellow brown green blue purple black	red orange yellow brown green blue purple black	red orange yellow brown green blue purple black

Comments:

Chapter 1 Assessment Record (continued)

K ☐
1 ☐

Name: ______________________________ LEA Teacher: ______________________

MATH

date:	September 199__	January 199__	May 199__
Knows names of 6 basic shapes	○ △ ▭ ⬯ □ ◇	○ △ ▭ ⬯ □ ◇	○ △ ▭ ⬯ □ ◇
Sorts	Color ☐ Size ☐ Shape ☐	Color ☐ Size ☐ Shape ☐	Color ☐ Size ☐ Shape ☐
Rote counts	1 to ______	1 to ______	1 to ______
Counts objects to 9	correct to ______	correct to ______	correct to ______
Recognizes numerals 0-9	knows: 0 1 2 3 4 5 6 7 8 9	knows: 0 1 2 3 4 5 6 7 8 9	knows: 0 1 2 3 4 5 6 7 8 9
Sequences numerals 0-9	correct to ______	correct to ______	correct to ______
Writes numerals from memory	writes: 0 1 2 3 4 5 6 7 8 9 see attached	writes: 0 1 2 3 4 5 6 7 8 9 see attached	writes: 0 1 2 3 4 5 6 7 8 9 see attached
Recognizes coins	penny ☐ nickel ☐ dime ☐ quarter ☐	penny ☐ nickel ☐ dime ☐ quarter ☐	penny ☐ nickel ☐ dime ☐ quarter ☐
Knows ordinals 1st - 5th	knows: 1st 2nd 3rd 4th 5th	knows: 1st 2nd 3rd 4th 5th	knows: 1st 2nd 3rd 4th 5th

Comments:

Chapter 1

Assessment, K-1

Alachua County Chapter 1
Developed by:
R. Zivanov, C. Clark, C. Cohen,
J. Hiebsch, S. Taylor

First Name

Date ________________

Last Name

Date ________________

Numerals

Date ________________

Recognizes Capital Letters Out of Order

Say, "I am going to point to some capital letters. I want you to tell me the names of the letters you know. If you don't know the name of the letter, you can say 'pass' or 'I don't know.' Are you ready?"

J C M B H

F A L E G

K D I

Recognizes Capital Letters Out of Order

Say, "I am going to point to some capital letters. I want you to tell me the names of the letters you know. If you don't know the name of the letter, you can say 'pass' or 'I don't know.' Are you ready?"

R T Z N Q

V Y P W S

U X O

Recognizes Lower-Case Letters Out of Order

Say, "I am going to point to some lower-case letters. I want you to tell me the names of the letters you know. If you don't know the name of the letter, you can say 'pass' or 'I don't know.' Are you ready?"

k h d a g

c m i e b

l f j

Recognizes Lower-Case Letters Out of Order

Say, "I am going to point to some lower-case letters. I want you to tell me the names of the letters you know. If you don't know the name of the letter, you can say 'pass' or 'I don't know.' Are you ready?"

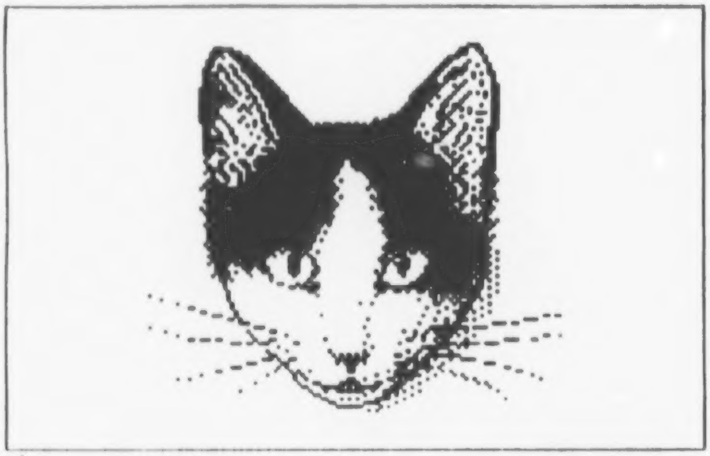
q v x n t

w o r z y

x u p

Letter-Sound Correspondence

Say, "I am going to show you a picture and tell you what it is. Cat. Listen. The word begins with a /**k**/ sound. The letter C makes that sound. I will name the picture. You say the word, then tell me the beginning sound you hear, and name the letter that makes that sound. Do you understand?"



BEST COPY AVAILABLE

Say, "Look at the butterfly. Tell me the letter that stands for the sound you hear at the beginning of butterfly." Teacher should review sample (cat) if necessary. Then begin by pointing out pictures at random, completing each picture page. If it is too difficult for the student, **STOP**.

Butterfly

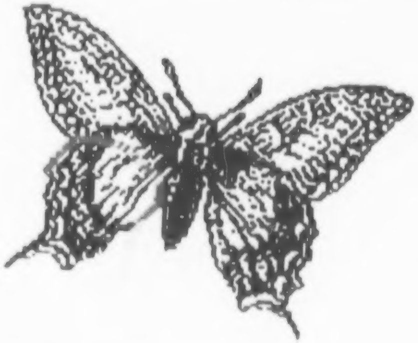
Cow

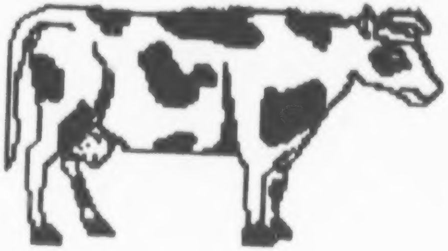
Dog

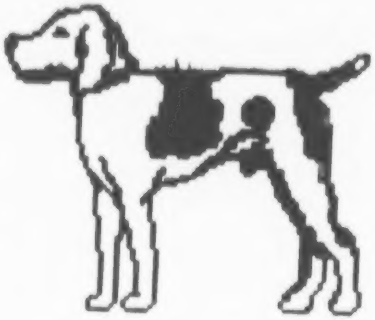
Fish

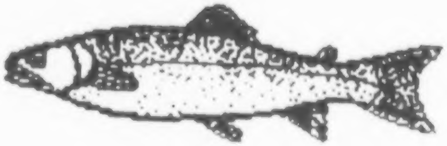
Gift

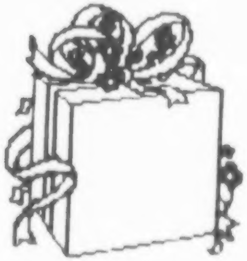
Horse

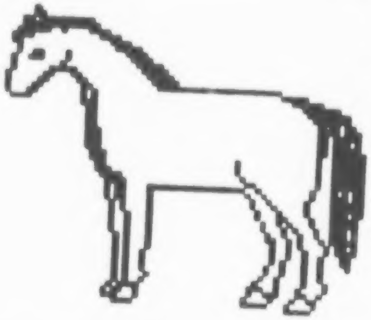






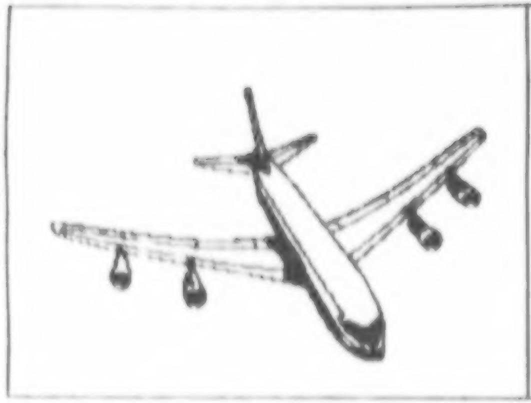


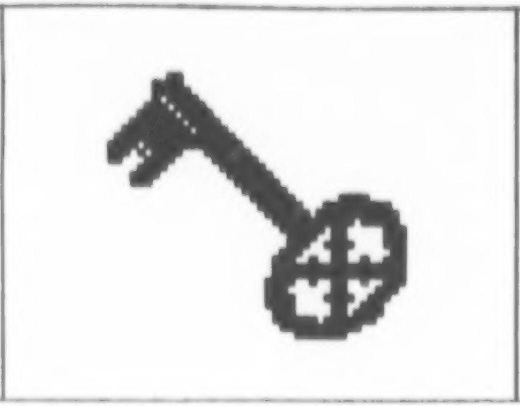


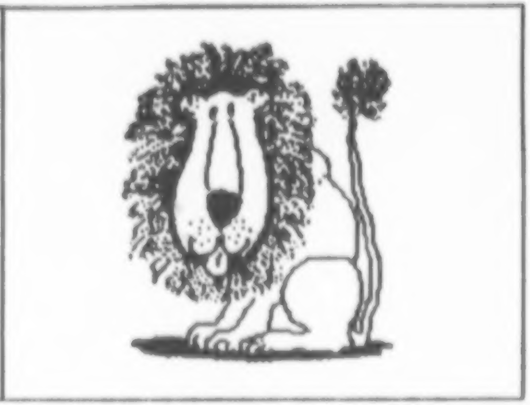


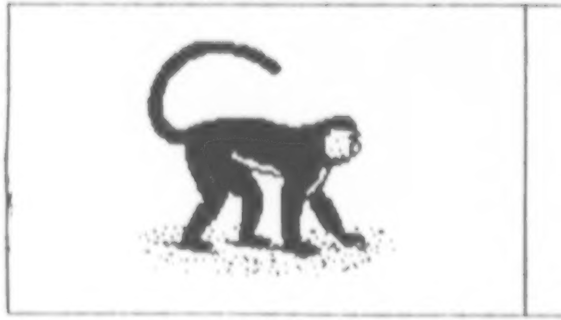
BEST COPY AVAILABLE

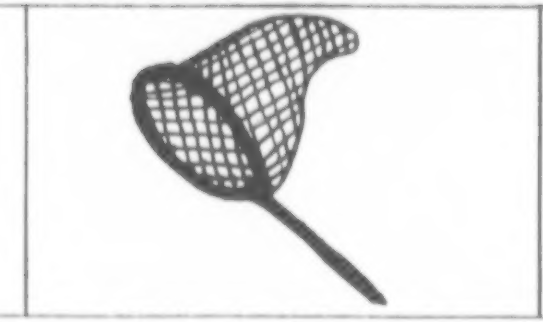
jet
key
lion
monkey
net
pig

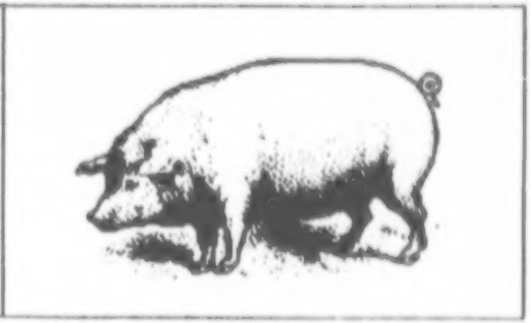






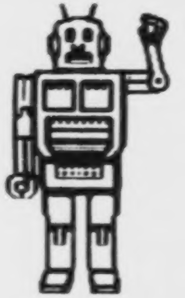


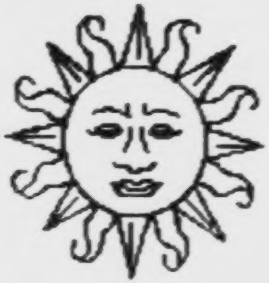


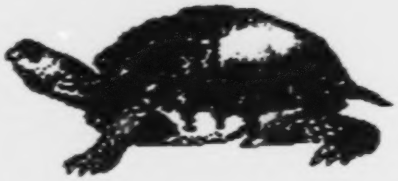


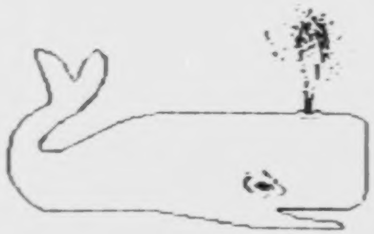
BEST COPY AVAILABLE

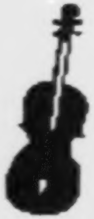
robot
sun
turtle
whale
violin
zipper

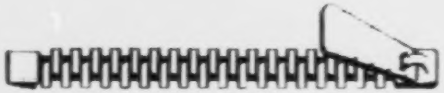












97

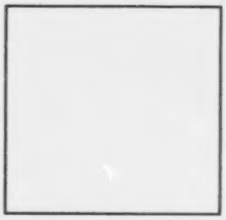
98

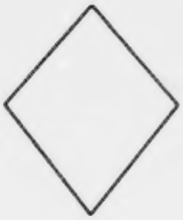
Colors

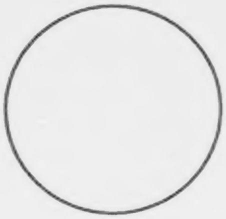
Say, "I am going to point to some colors. Tell me the names of the colors you know. If you don't know the name of the color, you can say 'pass' or 'I don't know.' What color is this?" (Teacher points)

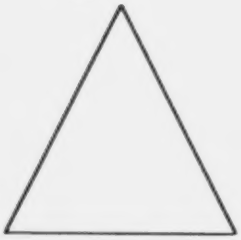
Shapes

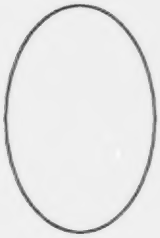
Say, "I am going to point to some shapes. Tell me the names of the shapes you know. If you don't know a shape's name, you can say 'pass' or 'I don't know.' Are you ready?"

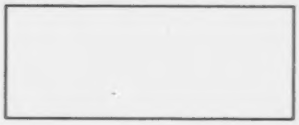












Sorts

Sorts

In order to observe the student sorting by color and shape, pattern blocks can be used. If the student can successfully sort the blocks by three different colors or shapes, record a check (✓) next to those items. For observing size sorting, use the Teddy Bear Family Counters and have the student sort by large, medium (or middle sized) and small.

Rote Counts

Rote Counts

Say, "I would like to hear you count as high as you can." If the student needs prompting, say, "Start with one. What comes next?"

Counts
Objects

Counts Objects

Teacher puts nine objects on the table. Say, "Count these for me." Allow time for counting. Ask, "How many did you count?"

Recognizes Numerals

Say, "I will point to some numerals (numbers). Tell me the names of the ones you know. If you don't know the name of the numerals, you can say 'pass' or 'I don't know.' Are you ready?"

8 5 2 9 6

3 7 1 0 4

Sequences
Numbers

Sequences Numerals

Say, "Here are some number cards. I want you to put them in counting order. How would you begin? What's next?"

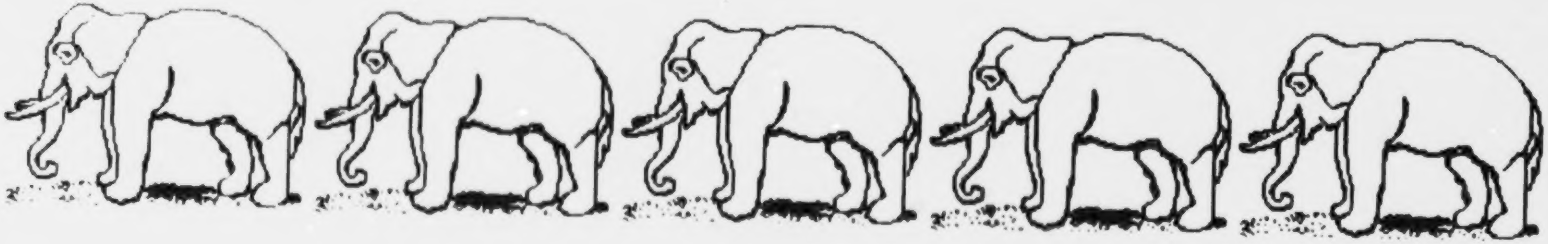
Recognizes Coins

Recognizes Coins

Say, "Here are some coins. Can you tell me the names of each?" (If the student knows the names, the student can be asked for further information. "Can you tell me how many cents a nickel (dime, quarter) is worth?")

Knows Ordinals

Say, "Here is a row of elephants. Can you point to the first elephant? Can you point to the fourth one? Point to the second elephant. Can you show me the fifth one? Point to the third elephant. Which one is the last in line?"



APPENDIX B

PREPARING FOR PORTFOLIO ASSESSMENT
(Extracted from Arter & Spandel, 1991)

For your portfolio system, who will be involved in planning? Who will have primary control over the decisions to be made? What leeway will there be for experimentation?

Which of the following purposes are of particular importance for the portfolio system you are developing?

- ☐ To show growth or change over time
- ☐ To show the process by which work is done as well as the final product
- ☐ To create collections of favorite or personally important work
- ☐ To trace the evolution of one or more projects/products
- ☐ To prepare a sample of best work for employment or college admission
- ☐ To document achievement for alternative credit for coursework
- ☐ To place students in the most appropriate course
- ☐ To communicate with student's subsequent teacher
- ☐ To review curriculum or instruction
- ☐ Large-scale assessment
- ☐ Program evaluation
- ☐ Other:

What are two major instructional goals for your program?

1.

2.

How will portfolios be used for classroom instruction/assessment in the system you are designing? What problems (if any) do you anticipate? What issues need to be resolved?

What questions would you consider asking students in order to prompt them to self-reflect on the work they are choosing for their portfolios?

What is the general curricular focus of the portfolio system you are planning?

☐ Reading
☐ Math
☐ Writing
☐ Integrated Language Arts
☐ Science
☐ Social Studies
☐ Fine Arts
☐ Other:

Keeping in mind the classroom goals for students you listed in #3, consider the kinds of things that might go into the portfolios you are designing in order to promote the attainment of those goals and, at the same time, provide good evidence of the achievement of those goals. First, what might be *required* to be included in *all* portfolios, if anything?

Second, list four categories of things that should be included in the work students select for their portfolios. How many samples of each of these things should students select?

1.

2.

3.

4.

Will you allow open-ended choices for the portfolio? How many open-ended items will be allowed?

Who will you get to assist you in finalizing these decisions?

What requirements will you have for when entries are selected for the portfolio, if any?

For your portfolio system you are developing, choose one of the types of products that students will be asked to place in their portfolio. What should a good performance look like? What does a poor performance look like? In other words, what are your criteria for judging performance?

For your portfolio system, which of the following considerations do you think are likely to be important in assessing the portfolio as a whole product?

☐ Amount of information included
☐ Quality of individual pieces
☐ Variety in the kinds of things included
☐ Quality and depth of self-reflection
☐ Growth in performance, as indicated in products or materials included
☐ Apparent changes in attitude or behavior, as indicated on surveys, questionnaires, etc.
☐ Other:

What criteria will you use to assess the student metacognition or self-reflection in the portfolio?

☐ Thoroughness
☐ Accuracy
☐ Support of statements by pointing to specific aspects of the work
☐ Good synthesis of ideas
☐ Self-revelation
☐ Other:

Who will help develop/select/adapt the performance criteria?

☐ Students
☐ Teachers
☐ Curriculum experts
☐ Evaluation and assessment experts
☐ Other:

How will you ensure that your criteria reflect current thinking concerning good performance in the area(s) you choose?

In your portfolio system, who will select specific work samples for the portfolio?

☐ Student only
☐ Teacher only
☐ Student and teacher
☐ Other:

How will storage and transfer occur, if at all?

Who will have ownership of the portfolio?

☐ The student alone
☐ The teacher(s) alone
☐ The student and teacher(s) together
☐ The school at which the portfolio is created
☐ Parents
☐ The student and parents together
☐ The school at which the portfolio is currently stored and used
☐ Other:

Who will have access to the portfolios?

☐ The student aand teacher(s) who created it
☐ Any teacher who needs/wants information provided by that portfolio
☐ Counselors
☐ Anyone in the school where the portfolio is housed
☐ Anyone from the district who shares an interest in the student's educational welfare
☐ Parents
☐ Other(s):

Imagine that you are planning to initiate your portfolio system during the coming year. Which of the following types of inservice would be most helpful to you and others that will be involved?

☐ Overview of the philosophy/rationale for use of portfolios
☐ Practical hands-on workshop on designing/assembling portfolios
☐ Ideas for portfolio management (e.g., ownership, transfer, etc.)
☐ Training in sound assessment practices, including use of portfolios in assessment
☐ Training in how to teach students good self-reflection skills
☐ Content area training
☐ Other:

SELF-TEST

Your self-test is performance based. Evaluate your responses to the boxed question using the following criteria:

1. **Completeness**: Look for the following:

   a. Did you answer *all* the questions? If not, did you have a good reason for not doing so? If you were unable to answer any of the questions right now, do you have a plan for how you will go about answering the questions?

   b. How much would it take to "clean up" your comments if they were going to be used as a discussion piece for a district/teacher committee looking into portfolios?

   c. Did you jot down other issues that should be addressed in addition to those listed?

2. **Quality**: Look for the following:

   a. What would be the reaction of each of the following groups to your plan—teachers, district personnel, students, parents, the school board, others? Did you take their points of view into account? If not, did you note why?

   b. Does your plan promote good instruction? If teachers carried out your design all year, would their students have received a good education?

   c. Does your plan promote good assessment? If your design were carried out, would you have quality information that has avoided the assessment pitfalls?

   d. Is your plan practical?

   e. Is your plan flexible?

3. **Individuality**: Look for the following:

   a. Does your plan match the curriculum in your classroom or district?

   b. Do your ideas reflect your own personal concept of what a good portfolio can/should be?

REFERENCES

Arter, J. "Performance Assessment: What's Out There and How Useful Is it Really?" Paper presented at the annual meeting of AERA. Chicago. (1991)

Arter, J. and Spandel, V. "Using Portfolios of Student Work in Instruction and Assessment." Northwest Regional Educational Library. (1991)

Finch, F.L. "Issues in Educational Performance Evaluation." *Educational Performance Evaluation*, Chapter 1. (1991)

Flood, J. and Lapp, D. "Reporting Reading Progress: A Comparison Portfolio for Parents. *The Reading Teacher*. (March, 1989)

Frechtling, J.A. "Performance Assessment: Moonstruck or the Real Thing." *Measurement Issues and Practices*. (Winter, 1991)

Linn, R., Baker, E.L., and Dunbar, S.B. "Complex, Performance-Based Assessment: Expectations and Validation Criteria." *Education Researcher*. (v. 20, No. 8, pp 15-21)

Mathews, J.K, "From Computer Management to Portfolio Assessment." *The Reading Teacher*. (February, 1990)

Olson, M.W. "Portfolios: Education Tools." *Reading Psychology: An International Quarterly*. (12:81-84. 1991)

Rothman, R. "Rand Study Finds Serious Problems in Vermont Portfolio Program." Education Week. December 16, 1992.

Valencia, S. "Portfolio Assessment for Young Readers." *The Reading Teacher*. (May, 1991)

Addressing Theoretical and Practical Issues of Using Portfolio Assessment on a Large Scale in High School Settings

Willa Wolcott, Ph.D.
Office of Instructional Resources
1012 Turlington Hall
University of Florida
Gainesville, Florida 32611

Funded by a grant from the Florida Department of Education

The use of writing portfolios—of various types and for different purposes—is gaining in popularity at all educational levels. Portfolios comprise a valuable instructional tool for both teachers and students, especially in the area of writing. Because they are gathered over an extended period of time, allowing for student revision, collaboration, and thought, portfolios serve as an appropriate vehicle for dramatizing the writing process. More significantly, they encourage the students to reflect on their own growth as writers and to participate actively in critically assessing their own work.

In addition to being recognized for their instructional value, writing portfolios have increasingly been advocated as a meaningful assessment tool. In fact, Grant Wiggins (1989) suggests that portfolios comprise the model for authentic assessment, which he defines as "the performance of exemplary tasks" (p. 703). Yet not everyone endorses the use of portfolios for assessment. For example, the National Council of Teachers of English (NCTE) Commission on Composition warns about the "bureaucratization" that results in "central offices" both telling what contents must be included in the portfolios and assigning numerical scores to portfolios for comparative purposes; as a result, the student and the teacher are removed from the process (See NCTE Council-Grams, *Portfolio Assessment Newsletter*, Sept. 1991).

To determine the feasibility of using portfolio assessment on a large scale, I received a grant from the Florida Department of Education last year to conduct a pilot study of writing portfolio assessment with three English classes at a local high school.[1] One class was a twelfth grade Advanced Placement class; one, a twelfth grade regular class that included basic writers; and one, an eleventh grade regular class that included basic writers. The study lasted for six months, and the 61 students were informed about the project beforehand. During that period I met frequently with the three teachers, all of whom were highly experienced teachers and very knowledgeable about writing assessment. The project was truly collaborative: At the meetings I talked about the issues we needed to consider, and they, in turn, indicated how well certain procedures might work with their own students and with the skills they were emphasizing in the classroom. It was important that the portfolio study not interfere with the curriculum.

A major purpose of the study was to address the theoretical issues that had surfaced in a review of portfolio literature and a review of portfolio practice elsewhere. In addition to the overriding need to design portfolio programs beforehand (French, 1991), these issues included (1) the optimal degree of standardization, (2) the need for teacher and student participation, (3) the question of authenticity, (4) the suitability of different scoring approaches in terms of reliability and validity, and (5) logistical problems of time, storage, and identification of scoring contexts.

Standardization of Portfolio Contents

The issue of standardization of portfolios for large-scale assessment is a controversial one. On the one hand, many educators, including Paulson and Paulson (1991), believe that standardization restricts the individualization of portfolios that is their real strength. On the other hand, French (1991) and Meyer, Schuman, & Angello (1990) believe that some standardization is necessary if portfolio data are to be aggregated; otherwise, they note, there is no basis for comparability. In order to facilitate the evaluations of our portfolios, we imposed some standardization both on the number of entries to be included in the portfolios and on the types of entries to be submitted as well.

First, the teachers administered in early September to each of the three classes a common, in-class topic titled "A person who has influenced your life." The purpose of the in-class topic was not only to provide some standardization of writing assignments across participating classes, but also to ascertain what students were capable of doing in an impromptu, timed writing situation. Six months later all three classes wrote again on a common, in-class topic suggested by one of the teachers and titled "A time in your life when you felt special." Interestingly, even though some Advanced Placement students initially objected to having to write on such generic, "bland" topics, their teacher reported that many subsequently became interested in the topics and in doing a good job about writing about them.

In addition to the in-class selections, students were asked to submit three other selections for their portfolios, including a best selection. They were also asked to provide a reflective letter in which they explained to the portfolio reader the reasons for their choice of the best selection. Finally, they were asked to submit a cover letter or form which gave background information about each paper, as well as the drafts that one paper had gone through. Thus, the ideal portfolio contained six selections in addition to a form providing the background of each entry and rough drafts of one piece.

---

[1] I thank teachers Gail Kanipe, Wendy McPhail, and Mary Morgan of Gainesville High for their wonderful participation in this project.

The types of selections to be included were broadly categorized to allow for the diversity of the differing curricula, to encourage the individuality of the students, and to foster the different types of writing taking place in the classroom. Thus, the first entry was supposed to be a "narrative or descriptive or informal essay," whereas the second entry was supposed to be a "persuasive or expository or academic essay." The third entry was designated the "best piece" and could take whatever form the student wanted it to. Having a comparable number of entries that addressed a comparable range of writing types would, we hoped, eliminate some of the problems encountered by raters in other portfolio programs, such as Vermont's, in which one scorer wrote that the meaning of a rating could differ substantially if it was based on a few, as opposed to many, selections.

As it turned out, discrepancies still arose. A number of portfolios were skimpy—either because students had not been present to write on one or both of the in-class writings, or because others chose as their "best" work a piece they had already selected to fit either the informal or formal category.

The portfolios of the Advanced Placement students, in particular, were rich and deeply textured, containing thoughtful academic papers on such topics as Shakespeare, the poetry of Donne, or "Hell in the Writings of Milton and Dante." Often for their "best" piece, the AP students chose a creative story, a poem or a play they had written. Even though the portfolios of the regular students typically did not contain as many academic essays as the AP students' portfolios, strengths appeared in the portfolios of some regular students as well. Indeed, their academic essays on such topics as "Heroism in Beowolf," "Macbeth," or "The Need for AIDS Testing," when juxtaposed against the informal writings of the students, allowed readers to see where the individual student's strengths or weaknesses lay.

As can be seen then, designating the number of entries and the broad types of writing to be included proved helpful in providing some comparable basis for evaluating the portfolios—even though unevenness continued to occur in the portfolios.

Scoring Procedures

Other central issues of portfolio assessment deal with scoring methods. One decision involves whether the entries will be scored individually or whether the portfolio will be evaluated in its entirety; in a number of programs, for example, such as those of Miami University of Ohio and the public schools of Cincinnati, Ohio, portfolios are scored as a whole. However, other educators, such as Peter Elbow (1991) or Richard Larson (1991), argue that the complexity of portfolios belies the giving of a single holistic or summative score. Another decision entails who will do the actual scoring—internal scorers who have the students in their own classes or external scorers who have had no real contact with the writers of the portfolios. For example, the practice followed by Vermont and Kentucky is to have teachers evaluate their own students' portfolios and then to have five of the portfolios selected at random and sent to an external committee for an independent scoring as verification. In several college programs in which high stakes are involved for the individual students, the evaluators have had no previous contact with the portfolio writers.

For the pilot study, eight writing instructors—including the participating teachers and myself—gathered one weekend to score the portfolios analytically and then holistically. I initially chose analytic scoring as the primary method because I felt that a single, holistic score might be difficult to assign in view of the variety of discourse forms contained in the portfolios, and I also wished to provide feedback to the students participating in the study. To do the analytic evaluations, the scorers used the scoring sheet (See Table 1) to rate each entry in the portfolios on nine different writing elements that dealt with rhetorical issues and with grammatical and mechanical concerns. One item measured the extent to which the reflective letter showed self-assessment skills, while an optional "bonus" category also enabled the scorers to reward exceptional creativity, voice, originality, or humor in the portfolios. Each element was rated on a scale of 1-4, with 4 being the highest. Prior to the actual scoring, the scorers trained together using two portfolios from the two regular classes. After reviewing written descriptors, the scorers each rated the portfolios independently and compared results, discussing their interpretations of the key whenever conflicts arose. The teachers participating in the pilot study generally refrained from scoring any of the portfolios written by their own students, because they felt that they lacked the necessary objectivity and tended even to become somewhat critical of their own students.

Scoring the portfolios analytically took an average of 15-20 minutes, with some portfolios, such as those from the Advanced Placement students, taking well over 30 minutes apiece. Table 2 depicts the results of the analytic scoring. Not unexpectedly, the AP class received consistently high average scores, with every portfolio receiving an average score of at least a 3 and several nearly achieving a perfect score of 4. A substantial range occurred for the students in classroom Z, with some students averaging over a 3 and some averaging below a 2. A range also occurred for students in classroom Y. This range is not surprising in view of the presence of some basic writers in regular classes. That no student's average score fell below a 1.5 could be significant in showing the value of revising as a tool that can help even with the weakest writers to improve; however, it must be noted that no penalties were assigned in this scoring for what might be missing from a given portfolio. That is, students' scores were averaged on the basis of what they actually submitted, rather than on the basis of what they should have included.

As a gauge of interrater reliability, 19 (30 percent) of the portfolios were given a second analytic scoring on the following day. These portfolios were selected at random, with at least five portfolios coming from each classroom. When alphas were run on the average score that each of the two readers gave on the sample of portfolios scored twice, the coefficient alpha was .83, denoting a reasonable interrater reliability rate for analytic scorings.

Most of the differences arose from contiguous scores. In 9 of the 19 portfolios the differences were consistently higher in both the rhetorical and the grammatical areas for one of the two readers, suggesting that in each case, one reader may have had the tendency to score more leniently or more harshly than the reader against whom she or he was paired. Still another contributing factor to the contiguous scores was the use of bonus points, which only one of any given two readers assigned in 5 of the 19 portfolios. In 11 of the 19 portfolios, splits—or non-adjacent scores—occurred on a few of the 46 specific items within the portfolios. Splits occurred **across** more portfolios on the rhetorical items, especially those items addressing thesis, focus, and thoughtfulness of content. However, the total number of splits **within** any given portfolio tended to be higher in the area of mechanics and grammar.

On the second day the scorers evaluated some portfolios holistically; the scorers did not rate holistically any of the portfolios they had previously scored analytically. Using a scale of 1 to 4 with 4 being the highest, the scorers assigned a single summative score that reflected the overall quality of the portfolios. Then, after giving the single score, they rated the overall quality of such key elements within the portfolios as development, content, sentence structure, and mechanics. They could also mark bonus categories for creativity and voice, and they could, if desired, add an optional comment. An important reason for including the individual ratings of key elements of the overall portfolios was to provide some feedback to students, a feedback which is lacking in holistic scoring and which is one strength of an analytic scoring. (In fact, portfolio advocates Paulson and Paulson (1991) have recommended that some combination of holistic and analytic scoring be done.) Table 3 illustrates the holistic scoring sheet.

Despite the diversity of discourse forms reflected in the portfolios, the scorers were readily able to assign a single, holistic score to reflect the overall quality of the writing. When alphas were run to determine the rate of interrater reliability, the coefficient alpha was .826, a figure comparable to the coefficient alpha for the analytic scale. Scorers spent approximately 6 minutes per portfolio in the holistic scoring, although the thick portfolios of the Advanced Placement students required more time. Precisely because of the high quality of many of the AP portfolios, all the scorers agreed that a broader scale, such as a 6-point scale, would be necessary to reflect the range of writing they encountered.

The scorers experienced little difficulty in marking their overall impression of the individual elements in the portfolios, and several wrote optional comments on the score sheets. Several scorers suggested including such features as diction, grammar, and usage in the individual ratings.

In order to see how well the original analytic scores correlate with the holistic scores of the 19 portfolios, the Pearson Product Moment Correlation was run; the Spearman Rank Order Correlation was also run to see how comparable the two scoring approaches ranked student papers. The Pearson correlation was .77, and the Spearman rank order correlation was .71. While not high, both correlations seem reasonable given the small sample size of 19 and the compressed scoring scale for the holistic scores—in which basically on the papers selected at random from the analytic set only scores of 2 through 4 were given.

The in-class topics, both of which had drawn on personal experience, proved accessible to everyone. Although students in all three classes tended to choose similar subjects to discuss, the stronger writers in the three classes often provided a fuller context for the influential people or events they wrote about, and they also demonstrated more insight into the actual meaning the person or the event had on their lives. One teacher participating in the study noted that she found it both interesting and helpful to see what the students in the other two classes had done with the same topics.

In all three classes, over 50 percent of the students **showed** improvement in their analytic scores from the first in-class topic to the last. (See Table 4). Even though these figures suggest that the majority of students in all classes showed growth, such a conclusion must be cautiously made. That is, in-class writing—with its time restrictions and lack of resources—negates much of the emphasis most writing classes put on revising, on multiple drafting, and on collaborative learning. Moreover, as the students who wrote on only one in-class topic were eliminated from analysis, the extent of any growth to be noted is limited to a portion of the pilot group as a whole. At the same time, it is encouraging to note that all three classes improved in the rhetorical areas, as well as in the area of mechanics and grammar. This finding counteracts the criticism of Knoblauch and Brannon that an emphasis on growth tends to focus on minor and measurable skills rather than on less measurable traits. (See Sommers, 1991)

Authenticating Student Writings

The in-class writings also served to authenticate the extent to which students have actually composed their own portfolios, even though authorship was never of real concern in this study since high stakes were not involved. However, this issue is a potentially troublesome one, especially in those situations in which students do have a lot at stake. Other means for authenticating ownership in this pilot study were the reflective letters and the multiple drafts that were required for one portfolio entry.

Reflective Letters

The reflective letters in the portfolios were scored differently from the other entries; that is, they were rated on the degree to which they reflected insight on the individual student's part into their own writing capabilities and performances. Not unexpectedly, two-thirds of the students in the Advanced Placement class received the highest scores possible for the insight their self-evaluations revealed. The analyses of these students often underscored their thoughtfulness, creativity, personal insight, and occasional, whimsical humor. Student 14X chose a serious work as an example of his best work for the following reasons:

> I chose "Elements in Shakespearean Comedy" as my best piece because I felt that it was my best analytical piece. I feel that I presented a clear thesis, developed it and proved it with parts from the play. . . This paper is much better than some that I wrote at the beginning of the year, which were unclear and unorganized. This piece reflects my improvement in those two areas and in understanding Shakespeare. These are some syntax and documentation errors but I have learned from those mistakes.
>
> *(Rating of 4)*

For some of the students in the regular classes, the reflective letters seemed to be difficult to write, and one of the teachers observed that a few of her students did not seem at all interested in the "why" of their choices. Nevertheless, in both regular classes, over half the students received at least upper-half scores for their self-assessment skills. For many of the students in these classes, their "best" piece was their "favorite" piece. Thus, a number of students talked honestly about the problems in certain papers which they nevertheless rated as their best for personal reasons or because they liked the topic. The reflective letter of student 7Y is typical of many such students' self-assessments.

> After looking through all of my writings, I chose the one that I thought was my best. It was a hard decision but the one I chose is the best example of my writing ability. The piece that I chose is about Benjamin Franklin's virtues, and how the world would be if everyone followed through on his virtues and used them as guidelines for their lives. I think this is my best writing because I demonstrate my ability to linger and link paragraphs and the way I express by beliefs and ideas and opinions. There's always room for improvement in my writings, but I thought that this one was the best example of me as a writer.
>
> *(Rating of 3)*

Despite the difficulties students seemingly experienced in writing the reflective letters, the self-evaluation such letters necessitated remain an important part of portfolio assessment. This study showed, as has the research (see Camp and Levine, 1991; Howard, 1990), that such awareness is not readily developed; hence, students need to be given several opportunities to reflect about their writing and to determine why some selections are better than others.

Logistics

The study also suggested that the logistics of portfolio collection and scoring should be standardized and simplified as much as possible. For example, requiring students to use a common cover form and to label the kind of entry that each submission represents would eliminate a potential source of confusion for the readers. In addition to making the portfolio entries easier to score, cover letters that explain the context of each submission also provide a fuller picture of each selection and thereby serve to authenticate the authorship of the portfolios. As Gentile (1992) points out, teacher notes explaining the background of the assignments are also helpful. Furthermore, to protect the privacy of students and teachers alike, the students' names on the portfolios should be masked and coded prior to a scoring, and teachers' grades or summative comments should also be covered.

The portfolios that appeared in the pilot study provided a good, in-depth picture of students' writing—of their strengths and weaknesses, their struggles and potential, and the progress they had made during the term. The portfolios revealed, moreover, the students as individual people.

The variety of abilities, discourse forms, and topics that were reflected in the portfolios did not present an insurmountable challenge for the scorers. Rather, the results suggested that scorers were able to assess the quality of the writings quite reliably given the complexity of the task. But what must be stressed is that the scorers, all of whom were highly experienced at scoring, underwent training for these specific scoring tasks.

Writing portfolio assessment on a large scale thus seems to be feasible provided that teachers are involved throughout the entire process and that safeguards are used to allow for individual creativity, for varying curricula, and for common requirements. Admittedly, the pilot study was small and included only three classes of varying abilities in one school, all of which had participated in the Florida Writing Enhancement Program. However, both the potential strengths and the potential weaknesses of portfolio assessment that surfaced in this study seem relevant to a number of other school systems.

Two key factors necessary for making portfolio assessment feasible are balance and participant involvement. That is, outside writings need to be balanced with some in-class work, just as informal writing assignments need to be balanced with academic essays. Standardized requirements need to be balanced with opportunities for individual writings, and student occasions for self-selection need to be balanced with guidance from the teacher. Finally, even though holistic scoring seems to be the most effective means of evaluating portfolios, that approach needs to be balanced with some analytic feedback to students or to schools. Again as the literature review notes (see Rigney, 1993) and as practice has shown, training in scoring portfolios is essential.

In addition to providing balance, a portfolio assessment procedure needs to ensure the involvement of students and teachers alike throughout the process. Portfolios imposed from the outside run the risk of being perceived as a time-consuming burden on everyone in the classroom (Cooper, 1991). However, portfolios that involve the teachers, as well as the students, in all phases of the portfolio program—from determining the nature and number of entries, to devising guides for reflective questions or generating descriptions of score levels—can become, as this pilot study suggested, a learning experience for everyone. When, as Camp suggests, portfolios are seen as an opportunity to demonstrate not only what has been learned but also what remains to be learned, then portfolios will have strengthened what should be an integral link between assessment and instruction.

REFERENCES

Camp, R. & Levine, D.S. (1991). Portfolios evolving: Background and variations in sixth through twelfth grade classrooms. In P. Belanoff and M. Dickson (Eds.), *Portfolios: Process and product* (pp. 194-205). Portsmouth, NH: Boynton/Cook.

Camp, R. (1990). Thinking together abut portfolios. *The Quarterly of the National Writing Project and the Center for the Study of Writing, 12*(2), 8-14, 27.

Cooper, W. (Fall 1990). Editorial, *Portfolio News, 2* (1).

DeWitt, K. (April 24, 1991). Vermont gauges learning by what's in portfolios. *The New York Times Education.* p. A23.

Elbow, P. & Belanoff, P. (1991). State University of New York at Stony Brook portfolio-based evaluation program. In P. Belanoff and M. Dickson (Eds), *Portfolios: Process and product* (pp. 3-16). Portsmouth, NH: Boynton/Cook.

French, R. (1991). *Issues and uses of student portfolios in program assessment.* A paper presented as part of a "Symposium examining assessment strategies of the Next Century Schools Project" at the AERA, Chicago, IL.

Gentile, C. (1992). *Exploring new methods for collecting students' school-based writing: NAEP's 1990 portfolio study.* Princeton, NJ: Educational Testing Service.

Howard, K. (1990). Making the writing portfolio real. *The Quarterly of the National Writing Project and the Center for the Study of Writing, 12 (2),* 4-7, 27.

Koratz, D. (January 1993). New report on Vermont portfolio project documents challenges. *National Council on Measurement in Education Quarterly Newsletter, 1* (4).

Larson, R.L. (1991). Using portfolios in the assessment of writing in the academic disciplines. In P. Belanoff and M. Dickson (Eds.), *Portfolios: Process and product* (pp. 137-150). Portsmouth, NH: Boynton/Cook.

Meyer, C., Schuman, S. & Angello, N. (1990). NWEA White Paper on *Aggregating portfolio data.* Lake Oswego, OR: Northwest Evaluation Association.

NCTE Council-Grams (September 1991). Portfolio assessment: Will misuse kill a good idea? *Portfolio Assessment Newsletter, 3* (1).

Paulson, F. & Paulson, P. (May 1991). *The ins and outs of using portfolios to assess performance.* An expanded version of a paper presented at the National Council on Measurement in Education in Chicago, IL.

Rigney, S. (January 1993). Vermont responds. *National Council on Measurement in Education Quarterly Newsletter, 1* (4).

Sommers, J. (1991). Bringing practice in line with theory: Using portfolio grading in the composition classroom. In P. Belanoff and M. Dickson (Eds.), *Portfolios: Process and product.* Portsmouth, NH: Boynton/Cook.

"This is my best. The report of Vermont's Writing Assessment Program (pilot year 1990-1991). Montpelier, VT. Vermont Department of Education.

Wiggins, G. (May 1989). A true test: Toward more authentic and equitable assessment. *Phi Delta Kappan.* 703-713.

Table 1
Scoring Sheet for Portfolios

Student I.D. ____________________

RATER NUMBER _____

	Narrative or Descriptive or Informal Essay	Persuasive or Expository or Academic Essay	Best Piece	First In-Class Writing	Second In-Class Writing	Reflective Letter	Total Score
1. The paper reflects either a clear purpose or a clear thesis that is stated or implied.							
2. The paper seems focused and organized.							
3. The paper is fully developed.							
4. The paper reflects a thoughtfulness of content.							
5. The word choice is appropriate for the subject at the secondary level.							
6. The sentence style is clear and varied, with appropriate sophistication for the secondary level.							
7. The paper reflects control of sentence structures. (Fragments, run-ons, and tangled syntax are avoided.)							
8. The paper reflects control of usage. (Errors in subject/verb agreement, pronouns, and dialect are avoided.)							
9. The paper reflects overall control of punctuation and spelling.							
10. The reflective letter shows self-assessment skills.							
BONUS: The ____________________ (creativity, originality, humor, voice) in this portfolio is noteworthy.							
OVERALL WRITING SCORE							

KEY

Very much so	= 4
To some degree	= 3
Not very much	= 2
Not at all	= 1
Not available	= NA

	Yes	No	Somewhat
Knowledge of the writing process is reflected in the drafts of one paper			

Sum of all papers _____

Highest score possible = 188
Lowest score possible = 46

BEST COPY AVAILABLE

Table 2.
Students' Average Analytic Scores

	Class X	Class Y	Class Z
Scores from 3.5 - 4.0	13 students (72%)		1 student (4%)
Scores from 3.0 - 3.4	5 students (28%)	2 students (11%)	4 students (17%)
Scores from 2.5 - 2.9		8 students (44%)	14 students (58%)
Scores from 2.0 - 2.4		7 students (39%)	3 students (13%)
Scores from 1.5 - 1.9		1 student (5%)	2 students (8%)
Scores from 1.0 - 1.5			

Table 3.
Prototype Holistic Scoring Sheet

Student I.D. ____________

RATER NUMBER: _____

OVERALL PORTFOLIO SCORE ☐

OVERALL RATINGS OF FEATURES

	Excellent	Very Good	Good	Fair	Below Average	Poor
Thoughtfulness of Content						
Development /Organization						
Diction						
Sentence Structure/Style						
Grammar and Usage						
Mechanics						
Self-Evaluation Skills						
OPTIONAL Creativity Voice						

OPTIONAL:
General Comments: ____

____

Table 4.
Students' Performance on the In-Class, Pre-Post Topics

Class	Number of Students who took Pre and Post Topics	Number of Students who Improved	Average Number of Points Improved	Number of Students who Declined	Average Number of Points Declined	No Change
X	14	8 (57%)	4.3 points	3 (21%)	3.7 points	3*
Z	17	9 (53%)	3.6 points	6 (33%)	2.2 points	2 (22%)
Y	13	7 (54%)	5.6 points	5 (38%)	2.6 points	1 (8%)

*Two students with perfect analytic scores on the pre-essay encountered a ceiling effect

The Effective use of Portfolio Assessment Within Preservice Teacher Education: The University of Florida's Elementary Proteach Program

Lynn Hartle
Paula DeHart

The Effective Use of Portfolio Assessment Within Preservice Teacher Education: The University of Florida's Elementary PROTEACH Program

Introduction

The University of Florida's elementary preservice teacher education program (PROTEACH) provides preservice teachers with opportunities to construct their knowledge of teaching through an active process that includes interaction with peers, professors, and cooperating teachers. Assessing each preservice teacher's performance requires collecting evidence of their progress throughout the program, both in coursework and field experiences. It is the belief of PROTEACH faculty that the preservice teachers themselves should be included in the organization and selection of evidence that they have critically analyzed and that they believe demonstrates their teaching competence. In accordance with PROTEACH program goals, the preservice teachers must show that they can reflect on their teaching practices and must demonstrate that they can make decisions about teaching and learning that are educationally sound and ethically defensible (Ross, Bondy, Kyle, 1993). The purpose of this paper is to present a newly developed constructivist assessment plan for preservice teachers, which is currently being field-tested through the collaborative efforts of field supervisors, preservice teachers, University of Florida faculty members, and Alachua County cooperating teachers.

This paper opens with an overview of the theoretical issues and research that formed the original creation and development of the PROTEACH elementary teacher education program and which continue to guide ongoing evaluation and revision of the program. Explained in the paper next, are the procedural steps taken by a University of Florida faculty committee to develop an assessment plan that was consistent with the program's constructivist focus and could provide a way to document the ongoing process of preservice teachers as they completed their program of study. As part of the development process, a sample group of preservice teachers, field supervisors and cooperating teachers were asked to field-test the new assessment plan and provide feedback to the faculty committee about their reactions to the new plan. Findings from the pilot study are included in the third section of this paper. In the concluding section of this paper strengths of the new constructivist assessment plan are discussed and recommendations for the plan's future use are proposed.

Background of Elementary PROTEACH

The University of Florida's elementary PROTEACH program strives to ensure that its graduates receive a comprehensive program of study based on a constructivist model of teaching and learning. Within this context, the term "constructivist" refers to the belief that learners construct knowledge through frequent interactions with their peers and the teacher, through exposure to a wide variety of curricular materials, and through the meaningful application of schooling to real-life situations (Ross, Bondy, Kyle, 1993). Central to the constructivist model of teaching and learning is the idea that students play an active role in their learning. This approach to education runs contrary to the more traditional "transmission model," which relies primarily on teacher lectures, textbook readings and prescribed (at the district or state level) curriculum that is taught to passive, receptive learners.

In order to help students construct knowledge, teachers need the skills necessary to transform the content of what is to be taught into meaningful and engaging learning activities. Teachers should have the tools necessary to assess their students' levels of understanding and to translate their assessment of student understanding into a curriculum that is appropriate for all students. These understandings need to be assessed in light of educational practices that are both educationally sound and ethically defensible, meaning that the practices not only result in student learning, but also are just and equitable (Ross, Bondy, Kyle, 1993). The complex tasks of evaluating student understanding and developing an equitable curriculum that facilitates student construction of knowledge requires that teachers use sound professional judgement and have the ability to make good decisions (Stone, 1987). In other words, within the constructivist model of teaching and learning, the teacher has the important role of being a "decision maker in a pluralistic society" (Geiger & Shugarman, 1988, p. 31).

Preparing teachers to be decision makers requires a much different approach to preservice teacher education than preparing teachers to transmit curriculum that has been predetermined by textbook publishers and state departments of public instruction. In keeping with the program goal of preparing teachers to be decision makers, professors in the elementary PROTEACH program make certain commitments in their own teaching. First, each professor makes an ethical commitment to empowering preservice teachers. This is essential because preservice teachers who do not feel they have control over their educational destinies in their teacher preparation classes will have a hard time viewing themselves as capable individuals who can empower their own students as learners.

Second, each professor commits to a constructivist approach to teaching and learning in their own teaching. Because knowledge about teaching is constructed, each professor must seek to discover the preservice teachers' underlying thoughts, beliefs and background experience. This is important because what happens in every learning situation is influenced by what the preservice teacher thinks and has experienced, and the beliefs that each preservice teacher holds as baggage from the sum of prior experiences can set up barriers to effective learning.

The third commitment every PROTEACH professor must make is to emphasize to preservice teacher the importance of being a reflective practitioner. Because each teaching situation is unique in terms of the teacher involved and the students' need and prior experiences, university professors cannot prescribe the "right" action to take for any given educational encounter. Thus, the professors must encourage preservice teachers to continually analyze their practices in light of their aims for education and the needs of their students. The preservice teachers must be taught to reflect on their practices through observation of student responses to their teaching and to utilize their knowledge of research and theory to set up learning situations that scaffold each child in his/her construction of knowledge.

Preservice teachers in the elementary PROTEACH program take part in a variety of experiences to realize both the abilities and attitudes to be truly reflective (RBK, 1993). Within several of the PROTEACH classes, classroom experiences are designed to help preservice teachers examine their prior experiential (tacit) knowledge in light of the research and the views of peers and professors. Both assigned and self-selected readings on contemporary practices provide yet another dimension within which preservice teachers can explore their knowledge of the teaching and learning process In addition, role playing and exploring case studies of teaching situations (focal situations) provide the motivation for preservice teachers to examine elements of teaching that they may have once taken for granted (Farber & Armaline, 1992). Finally, certain core classes within the PROTEACH elementary education program emphasize the development of the "cycle of reflective teaching" as the primary goal.

Revision of Preservice Teacher Assessment

Overview

Although faculty were satisfied that PROTEACH courses emphasized the program's focus on reflective teaching and a constructivist perspective on teaching and learning, they felt more work was needed to provide a link between experiences in the college classrooms and experiences in the field - both prior to and during internship. Providing a link between college classes and field experiences requires that teacher educators (professors and field experience supervisors) work more closely with teacher in the field, which is a recommendation currently being made in teacher education reform literature (Raines, 1992 & Quisenberry, 1987; Stone, 1987).

The literature on teacher education reform recommends that preservice teachers be paired with expert teacher practitioners and skilled supervisors during field experiences who can help them, individually and collectively, to apply meaning and theory to practice (Vandsleright & Putnam, 1991). Yet, as field experience are currently structured, there is often a lack of clear understanding on the part of each group involved as to what is expected of them. Quisenberry (1987) found that cooperating teachers are often unsure of what to do in their work with preservice teachers. Quisenberry also found that when the teachers do act, the guidance they provide to preservice teachers is not extensive or consistent enough to move the preservice teachers towards greater understanding of how each learning situation can help children to construct knowledge. The preservice teachers, themselves, often have unrealistic expectations for field experiences and find themselves dealing with the more managerial and technical aspects of teaching without having the opportunity to take a broader view of their roles as teachers during their field experiences. And the third group involved, the university field experience supervisors, frequently find it difficult to know how to structure supervision that is beneficial for preservice teachers and makes connections between university coursework and classroom field experiences in the limited time the supervisors have to spend observing and conferencing with each preservice and cooperating teacher.

To address the problem of discontinuity between PROTEACH coursework and field experiences and to better support preservice teachers in the field, a group of university faculty members and graduate students convened to search for ways to bridge the gap between the expectations of the university and those of the public schools. As a first step, the committee (based on feedback from university and public school faculty) decided the elementary PROTEACH program needed a more authentic assessment plan to evaluate preservice teachers in the field. They felt an assessment tool was needed that would provide university faculty, field supervisors, principals, cooperating teachers and preservice teachers with a better understanding of what is meant by "good teaching" and a clearer picture of how each individual can be involved in assessing preservice teachers' progress. A tool was needed that would clearly illustrate the constructivist point of view and would place preservice teachers in the role of professionals who are empowered to make decisions. The faculty committee felt an assessment tool was needed that would provide a more well-developed picture of preservice teachers' development than the checklist that was currently being used for evaluation in the PROTEACH program.

Venturing into the unknown waters of alternative assessment is risky and requires careful consideration of several critical issues facing alternative assessment (Mehrens, 1992 & Worthen, 1993). First, because cooperating teachers have limited time available to learn about and complete assessment forms and university supervisors have limited time available to instruct cooperating teachers on the use of a new assessment plan, the plan must be concise and easy to use. Second, because the assessment plan is intended to track the construction of each preservice teacher's instructional knowledge base, the assessment plan needs to be ongoing and systematic. The design of the plan needs to closely follow instruction and record both qualitative and quantitative aspects of the preservice teachers' performance.

A third critical issue related to alternative assessment is that the tool must be reliable and valid. Program goals must be reflected in the assessment plan and any evaluative tool used as a part of the plan must highlight teacher behaviors that coincide with program goals. For example, based on the PROTEACH program goals, any assessment tool would need to reflect a constructivist approach to teaching and learning and provide for collaborative opportunities between professors and preservice educators and between cooperating teachers and preservice educators (Valerie-Gold, Olson & Deming, 1992). This collaboration would empower preservice teachers by including them in the assessment process. An additional benefit to including preservice teachers in the assessment process is that it encourages them to introspect and reflect on their teacher practices, which is another PROTEACH program goal.

The New Constructivist Assessment Plan

The new assessment plan that was developed by the PROTEACH program faculty committee is based on a constructivist approach to teaching and learning and consists of three important components. The first component of the assessment plan is called the "Assessment Worksheet." It contains a set of eleven constructivist guidelines, with a list of possible characteristics and a comments section for each guideline (See Appendix A). The eleven "Constructivist Guiding Principles" that make up the assessment worksheet were adapted from the Kentucky Beginning Teacher Program (Kyle, et al., 1993). These principles are most closely aligned with the focus of the coursework in elementary PROTEACH and reflect the belief that all children deserve the right to learn and that knowledge is constructed. The possible characteristics listed under each guideline are examples of classroom practices that demonstrate the guideline is being met. The purpose of the worksheet is to allow whomever is making the assessment (i.e., the cooperating teacher, the field supervisor or the preservice teacher) to note evidence of the characteristics as they are demonstrated in the classroom. Within the comments section, the assessor might note which characteristics were demonstrated, when they were demonstrated and any other critical information related to the achievement of that constructivist guideline.

The second component of the assessment tool is a more standardized rating scale referred to as the "Assessment Form." Like the worksheet, the form includes the eleven constructivist guidelines; but rather than indicators, each guideline is followed by a developmental scale. (See Appendix B) The way the scale is designed allows the person completing the assessment form to place an X anywhere on a continuum starting on one end with no evidence that the guideline is being demonstrated to the other end of the scale which signifies mastery of the guideline. The scale was designed as a continuum to show that becoming a constructivist teacher is a developmental process that takes place over time. Using an assessment scale that is developmental within a teacher education program allows preservice teachers and their supervisors to chart the professional growth of the preservice teachers throughout their time in the program. The assessment form also helps preservice teachers to assess their strengths and identify areas in which they need to focus more attention.

The third component of the assessment plan based on the constructivist guidelines is the professional portfolio. Portfolios were chosen as an important part of the assessment plan because, while no grades on a transcript can ever tell a school administrator about a teacher's abilities as an effective educator, a portfolio can provide a window on that teacher's accomplishments. Beyond the limited definition of a portfolio as a collection of work, the model for the new constructivist assessment plan comes closest to the definition of "portfolio" developed by a consortium of educators under the auspices of Northwest Evaluation Association (Arter & Spandel, 1992):

> [A portfolio is a] purposeful collection of student work that tells the story of the student's efforts, progress or achievement in (a) given area(s). This collection must include student participation in selection of portfolio content; the guidelines for selection; the criteria for judging merit; and evidence of student self-reflection. The portfolio can address the complexities of teaching like no other document (Wolf, 1991).

Within the new assessment plan, each preservice teacher must build a portfolio that tells the story of their professional growth based on the standards defined in the constructivist guidelines. The portfolio provides a means for preservice teachers to keep a record of ongoing development of their skills, abilities and appreciation systems. The use of professional portfolios for assessing preservice teachers is not a new idea, but the use of the constructivist guidelines to direct artifact collection and focus the assessment of teacher growth is quite unique. Utilizing the constructivist guidelines for portfolio development and evaluation is a great strength of the new assessment plan because the guidelines provide clearly defined standards by which to compare each preservice teacher's professional growth.

Expected Benefits of the New Assessment Plan

The faculty committee that drafted the new constructivist assessment plan hoped the plan would strengthen the evaluation of preservice teachers' teaching performance in three important ways. First, the committee believed that a major strength of the new assessment plan is that it provides clearly articulated goals and expectations for all persons involved in field experiences - principals, cooperating teachers, college faculty supervisors, and preservice teachers. The hope was that by clearly defining the goals and expectations of the elementary PROTEACH teacher education program all participants would be empowered. University supervisors would be better able to communicate with elementary school personnel about the program's goals and expectations for preservice teachers. Cooperating teachers would be better able to assist preservice teachers in meeting goals and expectations while at the same time assisting preservice teachers in the development of professional portfolios. Preservice teachers would have a more concise overview and review of what they had learned in coursework about good teaching principles and would be able to see connections between these teaching principles and their classroom practices.

A second, but equally important strength of the new assessment plan is that it actively involves preservice teachers in their own assessment. This new assessment plan provides opportunities for preservice teachers to select evidence of good teaching from coursework assignments and field experiences to include in their professional portfolios. The new assessment plan also encourages preservice teachers to use advice from college faculty, the current research on best practices for teachers, and their own experiences with children in schools to make professional judgments and evaluate their own progress on the assessment worksheet and form. An additional benefit of involving preservice teachers in self-assessment at the preservice level is that they continue to evaluate and review their own teaching practices throughout their teaching careers.

A third strength of the new constructivist assessment plan is that it prepares preservice teachers to meet the qualifications outlined for teachers in Goal #6 of Florida's "Blueprint 2000: A System of School Improvement and Accountability." Successfully demonstrating the skills and abilities defined in the new PROTEACH assessment plan, including the development of a professional portfolio, helps to ensure that preservice teachers also possess the professional skills required of Florida teachers. These include the abilities to use the following: the principles of continual quality improvement in an instructional setting with their students; appropriate strategies for teaching students from diverse cultural backgrounds, with different learning styles, and with special needs: and appropriate skills and strategies that promote the creative/critical thinking capabilities of students. By possessing these qualities, the preservice teachers will have the skills necessary to contribute to the effort to improve Florida's schools.

Pilot Study of the New Constructivist Assessment Plan

Purposes for Pilot Testing

Beginning in January, at the start of the University of Florida's spring semester, the new constructivist assessment plan was field-tested in several classrooms in Alachua County. The purpose of field-testing the new plan in a limited number of classrooms and collecting pilot data before widespread use was twofold. First, although the assessment plan was developed based on the latest educational research on teaching and assessment, an assessment plan based on constructivist guidelines had never been used before. Until the plan was field-tested, there was no guarantee it would work successfully with preservice teachers. Testing the assessment plan in a few classrooms provided the opportunity to validate the effectiveness of the plan and to work out any bugs that might arise with its use.

The second purpose for field-testing the new assessment plan was to elicit feedback from those participants who are responsible for implementing the plan and who ultimately determine its success or failure, that is the cooperating teachers, preservice teachers, and university faculty and supervisors. During the field-testing process it was explained to all of the participants that the assessment plan was in a formative stage and that their feedback would be used to make revisions to strengthen and simplify the assessment process.

Because the new assessment plan was intended to be developmental and used throughout the PROTEACH program, it was determined the plan should be field-tested with preservice teachers at each of the field experience levels. Thus, pilot data were collected from university supervisors, cooperating teachers, and preservice teachers in each of the two pre-internship semesters, which occur in the second and third semesters of the PROTEACH program; and during the full-time internship, which occurs during the preservice teachers' fifth year. Data were collected through informal interviews with cooperating teachers, formal and informal interviews with preservice teachers, written responses on the assessment worksheets and forms, and notes kept by the university supervisors, who in this case are also the researchers. Preservice teacher progress was monitored by the university supervisor through weekly observations and was documented in observation logs. Preservice teacher progress was also monitored through weekly verbal communication between the university supervisor and cooperating teacher. Throughout the semester, the preservice and cooperating teachers were invited to actively participate in every step of the assessment process. (Because data collection is continuing, the data reported in this paper are limited to feedback from the supervisors/researchers, cooperating teachers and preservice teachers involved in full-time internships.)

Initial Feedback from Participants

Initial feedback from supervisors/researchers, preservice teachers and cooperating teachers about the new assessment plan provided valuable information about how the assessment worksheet and form were used. Participants' responses also provided insight into how the assessment worksheet and form, based on the constructivist guidelines either were or were not utilized by preservice teachers in their professional portfolio development. Analyses of data were accomplished utilizing a framework provided in a chapter titled, "Portfolio Assessment: A Panacea or Pandora's Box?" (Valencia, 1991). In the following paragraphs an analysis of pilot data will be reported and discussed, and strengths and problem areas of the constructivist assessment plan will be identified.

Strengths

The primary areas of strength of the new assessment plan as outlined by Valencia (1991) and identified in this study's pilot data were the authenticity and quality of the plan. The authenticity and quality of the assessment plan were reflected in two ways. First, the assessment plan, particularly the worksheet and form, were noted by participants to have a clear connection to PROTEACH program goals. None of the teachers, students or university supervisors noted discrepancies between what they knew the goals of the PROTEACH program to be and the constructivist guidelines outlined on the assessment worksheet and form. During formal interviews, most of the preservice teachers said they recognized the "Constructivist Guidelines" as those that were emphasized during their undergraduate and graduate coursework. One preservice teacher explained that when she first read through the assessment guidelines her reaction was, "Yes, that's what a teacher does," indicating that her understanding of what a teacher does, based at least partially on her experiences in the PROTEACH program, matched with the image of teaching delineated by the constructivist guidelines.

The authenticity and quality of the assessment plan were also reflected in preservice teachers' comments about how the constructivist guidelines helped them to focus what they did and how they reflected on their experiences in the classroom. Some of the comments made by preservice teachers in this area included references to how the constructivist guidelines helped them to think about taking advantage of teachable moments, collecting writing samples to evaluate student learning, getting parents involved in the classroom, and addressing issues of student diversity in their instruction. In each of these examples, the preservice teacher mentioned that these were important components of classroom instruction that they did not feel that they would have considered without their use of the constructivist assessment plan.

The third area of strength related to the authenticity and quality of the new assessment plan was reflected in the overwhelmingly positive response to the use of the plan by all participants. University supervisors said they were pleased with how closely the assessment plan matched program goals and with the formative nature of the plan. Because the assessment plan is formative rather than summative, the supervisors felt they could emphasize a constructivist perspective on learning to teach and stress that the construction of teacher knowledge occurs throughout a teacher's career. A formative approach to preservice teacher evaluation runs contrary to the summative approach typically used in teacher education programs where preservice teachers are evaluated at the end of their internships as to whether or not they have "made it" as teachers.

The preservice and cooperating teachers also responded favorably to the new assessment plan. There was concern on the part of the university faculty committee at the beginning of the semester that cooperating teachers might be put off by the additional time required to complete the new assessment worksheet and form, but this was not the case. The cooperating teachers were pleased that PROTEACH faculty were trying to improve the assessment process and clarify the program's goals and expectations for preservice teachers. One cooperating teacher commented that she was happy that the assessment plan reflected such high standards for teaching performance. The preservice teachers had suggestions for the revision and improvement of the assessment plan, but all recommended that the plan continue to be used in the future.

Although the pilot data reflected several strengths of the new assessment plan, there were also some problem areas identified. One such problem area was the difficulty preservice and cooperating teachers had making the shift from a summative to a formative evaluation tool. Even though the developmental nature of the assessment process was explained to cooperating and preservice teachers, formal and informal interviews revealed that they did not refer to the assessment worksheet or form until they were asked to submit the assessment forms to the university supervisors at the midpoint and end of the ten-week internship. Because they viewed the assessment as summative, the preservice teachers demonstrated little evidence that they used the plan to help them reflect on their ongoing professional development. Likewise, because the cooperating teachers had a hard time shifting to a formative assessment plan, they did little to assist the preservice teachers in ongoing assessment.

Another problem area that calls for attention is in the area of empowerment for preservice teachers. Each preservice teacher was given a copy of the assessment form and assessment worksheet with the intent that they would assess their own teaching and share their perspective with their cooperating teacher and university supervisor. In reality, there was little evidence in interviews and observations that the preservice teachers took an active role in evaluating their teaching or sharing their assessment of their teaching with others. Although preservice teachers' comments indicated that they felt it was important to think about their teaching, they still relied primarily on the cooperating teacher and university supervisor to tell them what they did well and what they needed to improve.

A third problem area identified during field-testing was that the preservice teachers did not seem to make a connection between the constructive guidelines provided in the assessment forms and the development of either instructional activities or their professional portfolios. The preservice teachers did not seem to view the constructivist guidelines as helpful in developing learning experiences for children or in gathering evidence for a portfolio that they could review regularly to note their own progress. Most of the preservice teachers' responses indicated that they knew they were to collect evidence of what they had learned about teaching, but that they saw the portfolio as a storage device to hold "things" that they would later take to a job interview.

All of the problem areas identified during field-testing of the new constructivist assessment plan indicate that there is a dire need for better staff development in the use of the plan. In order for the new plan to be used successfully and effectively, all the participants must understand the constructivist guidelines and must see the connections between the constructivist guidelines, the assessment worksheet and form, and the professional portfolio. While each cooperating teacher received a cover letter explaining the new assessment plan, the letter alone could not address the novelty of this form of assessment. Cooperating teachers did not realize how the worksheet could be used to document the preservice teachers' ongoing progress and did not take out the worksheet until the midterm and final evaluations were due.

According to field-test data, intensive staff development is also crucial for the effective use of the assessment plan by preservice teachers. Although the assessment plan was introduced and explained during an internship seminar, analyses of pilot data revealed that the preservice teachers did not understand that the plan was formative; they did not make connections between the constructivist guidelines defined in the assessment forms and either the planning of instructional activities or the development of a professional portfolio; and they did not feel empowered to actively participate in the assessment process. These findings indicate that several internship seminars may have to be fully or partially dedicated to discussion of the constructivist assessment plan and the preservice teachers' role in assessment. The findings also indicate that university supervisors need to closely monitor the preservice teachers' understanding and use of the assessment plan throughout the semester so that misunderstandings can be corrected before the semester ends.

Recommendations for Future Use of the PROTEACH Constructivist Assessment Plan

Based on the results of data gathered during field-testing, the PROTEACH faculty has decided to use the constructivist assessment plan, including the professional portfolio, for future PROTEACH preservice teacher assessment during all field experiences. Dialogue among faculty at the University of Florida has already begun regarding extensions of the use of the tool and portfolio guidelines in other aspects of the PROTEACH teacher education program (i.e., the guidelines may be distributed to preservice teachers in their first semester enrolled so the guidelines can be used as ongoing assessment over the entire three years enrolled in PROTEACH). Preliminary data have provided staff at the University of Florida and faculty in Alachua County schools with evidence that this new constructivist assessment plan merits continued use.

Data collected from this pilot study reveal that the assessment plan can be beneficial in the areas that were identified as important during the initial planning stage. Expected as the major strength, the constructivist guidelines clearly presented principles and characteristics of good teaching and learning consistent with the goals of the PROTEACH teacher education program. Preservice teachers, interviewed during their internships, recognized connections between these guidelines and what they had experienced during coursework in PROTEACH. Preliminary findings suggest the need for future research regarding preservice teachers' understandings of the comprehensive nature of constructivist-based education and their knowledge of how to plan lessons that are constructivist. Preliminary findings also indicate the need for increased staff development for preservice teachers to facilitate their reflection on and development of professional portfolios based on constructivist principles.

A second important strength of the new assessment plan is that it does actively involve preservice teachers in their own assessment. Although the results of the pilot study revealed that the preservice teachers did not become as involved in self-assessment as was hoped, comments made by the preservice teachers during interviews revealed that the constructivist guidelines did encourage them to think about aspects of their teaching they might not have considered otherwise. Further research on the assessment plan should focus on the extent and nature of staff development that is necessary to engage preservice teachers in ongoing assessment of their teaching practices using the constructivist assessment plan. Additional study should also focus on what can be done to encourage preservice teachers to see themselves as equal participants in a three-way discussion of their professional development.

And finally, there is little doubt among University of Florida professors that the PROTEACH preservice teacher education program provides educational experiences that address critical issues for teacher reform as outlined in Florida's "Blueprint 2000: A System of School Improvement and Accountability." PROTEACH coursework and the new constructivist assessment plan are designed to help preservice teachers develop professional skills required of Florida teachers. Examples of these important skills are the ability to appropriately teach children from diverse cultural backgrounds, with different learning styles, and with special needs, and the ability to use skills and strategies that promote the creative/critical thinking capabilities of students. These skills are essential for teachers who must prepare children to succeed in an increasingly diverse and complex society.

In conclusion, what this study serves to illustrate is the importance for teacher educators to continually search for ways to send consistent messages to preservice teachers about appropriate and effective teaching practices. These messages about good teaching are reinforced when they are reflected in coursework, in field experiences and, as this study indicates, in assessment techniques. In the University of Florida's PROTEACH program, faculty members want their graduates to leave with clear ideas about how to develop instruction that enables children to construct knowledge. It is also the hope of PROTEACH faculty that their students become reflective thinkers who make informed decisions based on research and theory, and can move beyond personal bias to provide all children equal access to educational opportunities.

REFERENCES

Arter, J.A. & Spandel, V. (1992). Using portfolios of student work in instruction and assessment. *Educational Measurement: Issues and Practices, 11*(1), 36-44.

Dittmer, A.E., Fischetti, J.C. & Kyle, D.W. (1993). Constructivist teaching and student empowerment: Educational equity through school reform. *Equity and Excellence in Education, 26*(1), 40-45.

Farber, K.S. & Armaline, W.D. (1992). Unlearning how to teach: Restructuring the teaching of pedagogy. *Teaching Education, 5*(1), 99-111.

Geiger, J. & Shugarman, S. (1993). Portfolios and case studies to evaluate teacher education students and programs. *Phi Delta Kappan. 74* (6), 31-34.

Knowles, J.G. & Holt-Reynolds, D. (1991). Shaping pedagogies through personal histories in preservice teacher education. *Teachers College Record. 93*(1), 87-113.

Mehrens, W.A. (1992). Using performance assessment for accountability purposes. *Educational Measurement: Issues and Practice. 11*(1), 3-9, 20.

Quisenberry, N.L. (1987). Teacher education: Challenge for the future. *Childhood Education, 63*(4), 243-246.

Raines, S.C. (1992). Promising research, practices and developments: Strengthening student teachers' reflections through collaborative research. *The Journal of Early Childhood Teacher Education, 13*(42), 14-15.

Raines, S.C. (1992). Promising research, practices and developments: Strengthening student teachers' reflections through collaborative research. *The Journal of Early Childhood Teacher Education, 13*(40), 19-21.

Ross, D., Bondy, E., & Kyle, D. (1993). *Reflective Teaching for Student Empowerment*. New York: Macmillan

Stone, B. (1987). Learning to teach: Improving teacher education. *Childhood Education. 63*(5), 370-377.

Surbeck, E. (1991). Assessing reflective responses in journals. *Educational Leadership, 48*(6), 25-27.

Valencia, S.W. (1991). Portfolios: Panacea or pandora's box? In Finch, F. (Ed.), *Educational Performance Assessment* (pp. 33-46). Seattle: Riverside Publishing.

Valeri-Gold, M., Olson, J.R. & Deming, M.P. 1993). Portfolios: Collaborative authentic assessment opportunities for college developmental learners. *Journal of Reading, 35*(4), 298-305.

Vansledright, B.A. & Putnam, J. (1991). Thought processes of student teachers. *Teaching & Teacher Education, 7*(1), 115-118.

Wolf, K. (1991). The schoolteacher's portfolio: Issues in design, implementation, and evaluation. *Phi Delta Kappan, 73*(2), 129-136.

Worthen, B.R. (1993). Critical issues that will determine the future of alternative assessment. *Phi Delta Kappan, 74*(6), 444-454.

Appendix A

Intern Assessment Worksheet

Cooperating Teacher's Name ______________________________
Pre-intern/Intern's Name ______________________________
Date ______________________________

Teaching Guidelines

1. The purpose of instruction is to actively involve each child in the construction of knowledge.

   Possible Characteristics:
   - Teacher demonstrating a sophisticated knowledge of content
   - Using writing, visual art, or performing art which stress student reflection
   - Providing assistance to students through reciprocal teaching, scaffolding, peer teaching, modeling, and demonstrating
   - Using multiple modes of performance such as videotape, painting, role plays, simulations, or presentations
   - Connecting learning with children's lives
   - Encouraging student-generated questions
   - Providing audiences beyond the teacher and the classroom for student work

2. The pre-intern/intern, for purposes of teaching and learning, creates a sense of student belonging and psychological safety.

   Possible Characteristics:
   - Designing ways for students to understand human commonalities and differences while honoring diversity.
   - Encouraging risk taking, divergent thinking, and tolerance of new ideas
   - Creating a classroom community emphasizing cooperation
   - Establishing classroom rule which stress student choice, responsibility, roles, and power
   - Helping children value the unique individuality of each child in the classroom
   - Teaching conflict resolution and communication skills
   - Understanding equity; modeling respect and empathy for class members
   - Building self-esteem and a healthy sense of humor

3. The pre-intern/intern is able to recognize decision points within lessons and their potential impact on student learning.

   Possible Characteristics:
   - Capitalizing on the "teachable moment"
   - Providing a rationale for teaching decisions to students
   - Making on-the-spot decisions grounded in instructional goals, knowledge of students, and total student context
   - Showing awareness of the "big picture," represented by lessons
   - Encouraging and responding to student questions

4. The focus of instruction is on lessons that are inherently meaningful to children.

   Possible Characteristics:
   - Developing lessons that enable children to use what they learn now
   - Taking time to get to know the students
   - Creating audiences beyond the classroom for the work of the class
   - Using manipulatives and multi-sensory approaches

Relating classroom learning to the real world
Using varied modes of exploration and performance

5. The teaching/learning process emphasizes teacher-student and student-student interaction and collaboration.

Possible Characteristics:
- Using cooperative learning; Establishing groups—tutoring, sharing, revising, responding, and assessing
- Implementing various aspects of cognitive apprenticeship such as modeling, demonstrations, and peer teaching
- Using simulations and games for understanding and team building

6. The pre-intern/intern demonstrates and uses knowledge of the students' culture, prior understanding, misconceptions, beliefs, values, etc., in planning instruction.

Possible Characteristics:
- Demonstrating and using knowledge of student culture to help students construct knowledge
- Making positive contacts with the home (e.g., home visits, phone calls, memos)
- Using non-prejudicial language and behavior; honoring uniqueness of children
- Requesting parental perspectives on children's performance
- Making instructional choices which connect the values and beliefs of children to school knowledge
- Initiating instruction with assessment of student understandings

7. The pre-intern/intern demonstrates an in-depth knowledge of content as well as the ability to create conceptual links across subject areas.

Possible Characteristics:
- Employing interdisciplinary units
- Using team teaching and professional teamwork
- Linking concepts across the disciplines and teaching thematically
- Connecting activities across grades, disciplines, and school themes
- Organizing information with minimum confusion
- Giving clear directions
- Developing a logical sequence for instruction

8. Pre-intern/intern provides an informed rationale for instructional decisions.

Possible Characteristics:
- Studying research and best practice and citing professional literature
- Attending professional meetings
- Identifying areas for personal and professional growth
- Exploring prior experience of students (interest inventories, journals, logs, diaries)
- Recognizing outmoded practices and beliefs
- Using research and theory as the basis for challenging assumptions
- Drawing up (and constantly testing) prior knowledge derived through practical experience

9. The pre-intern/intern continually analyzes evidence of student learning as a basis for instruction.

Possible Characteristics:
- Constructing portfolios
- Assessing writing samples
- Collecting multiple kinds of evidence of student learning
- Paying attention to student responses (particularly questions, not just answers) and helping children probe more deeply

Citing evidence of student learning in instructional decisions

10. The pre-intern/intern not only demonstrates technical competence in planning, implementing, and organizing instruction and classroom management but also organizes and manages the classroom in ways that foster student self-discipline.

Possible Characteristics:

Involving students in developing/ revising classroom rules and routines
Using descriptive rather than judgmental language in response to students who are "incorrect" in responses or "off-task"
Addressing student repeated inappropriate behavior through collaborative problem-solving approach
Communicating clear expectations for student performance in class activities
Providing clear instructions and explanations which acknowledge a variety of engagement styles in children
Using pacing which is appropriate for instructional activities

11. The pre-intern/intern demonstrates the ability to coordinate human and material resources to enhance student learning.

Possible Characteristics:

Providing varied opportunities for parent involvement in classroom activities
Engaging students in oral histories and intergenerational projects
Involving community consultants to assist students with in-class and out-of-class linking of experience and education
Using a variety of materials and employing them skillfully

APPENDIX B

Cooperating Teacher's Name ____________________
Pre-intern/Intern's Name ____________________
Date ____________________

Pre-Intern/Intern Assessment Form
SAMPLE PAGE

For each of the following guidelines, please place an X anywhere along the continuum to indicate to what degree your pre-intern/intern demonstrated characteristics of that guideline. You are encouraged to share this form and the assessment worksheet with your pre-intern/intern.

NE = no evidence of this guideline observed
B = beginning to demonstrate characteristics of this guideline
D = demonstrating some characteristics of this guideline and continuing to develop in this area
M = mastery of this guideline

1. The purpose of instruction is to actively involve each child in the construction of knowledge.

NE B D M

Comments:

2. The pre-intern/intern, for the purposes of teaching and learning, creates a sense of student belonging and psychological safety.

NE B D M

Comments:

3. The pre-intern/intern is able to recognize decision points within lessons and their potential impact on student learning.

NE B D M

Comments:

4. The focus of instruction is on lessons that are inherently meaningful to children.

NE B D M

Comments:

5. The teaching/learning process emphasizes teacher-student and student-student interaction and collaboration.

NE B D M

Comments:

6. The pre-intern/intern demonstrates and uses knowledge of the students' culture, prior understanding, misconceptions, beliefs, values, etc., in planning instruction.

NE B D M

Comments:

Portfolio Assessment in Teacher Education Courses

Lyn Wagner, Ann T. Agnew, and Dana R. Brock

Department of Elementary & Secondary Education
The University of West Florida
11000 University Parkway
Pensacola, FL 32514

Portfolio Assessment in Teacher Education Courses

We initiated the use of portfolio assessment in two different undergraduate teacher education courses we teach at the University of West Florida. We did so in the belief that students need to develop familiarity with forms of performance assessment that they may be expected to use in their own classrooms when they enter the profession. To develop familiarity, we believe that they should participate in portfolio assessment as one means for documenting their own learning.

Another reason for including portfolio assessment in the preservice curriculum was to support the theme of the teacher education program at UWF: The Teacher as an Empowered Professional. As students proceed through the program, we want them to participate in experiences that mark the Empowered Professional as one who engages in reflective thinking, learns from social interaction with professional peers, becomes an informed decision maker, and sets personal learning goals. Participation in portfolio assessment to document one's own learning provides numerous opportunities to engage in such empowering experiences.

A portfolio has been described as a container for the storage of evidence of an individual's knowledge and skills. Artists, for example, have traditionally used portfolios to document their achievements. However, limiting the description of a portfolio to a storage device fails to recognize that the use of portfolios implies a dynamic view of assessment. Portfolio use embodies the belief that learning is more richly and accurately portrayed when based upon multiple sources of evidence collected over time in contextually relevant settings (Paulsen, Paulsen & Meyer, 1991; Valencia, 1990; Wolf, 1991).

In this article, we describe our initial experiences of using portfolio assessment in two sections of a language arts methods course and a section of an early childhood curriculum course. We explain how course goals were used to define the purposes to be achieved by use of the portfolio strategy. We describe the steps we took to help students prepare portfolios and the management strategies that we found helpful. Finally, we explain how we evaluated student portfolios and how the portfolio evaluation fit into the evaluation scheme for each of the courses.

Use of Course Goals to Structure Portfolios

Course goals were used to determine the purposes to be served by use of portfolios. In the case of the language arts methods course, the course goal used to structure the portfolio project for the students can be stated as follows: **Each of you needs to further develop understanding of yourself as a language user.** For the early childhood course, the goal to be supported was: **Each of you needs to further develop understanding about how your personal and professional experiences influence your beliefs about learning and teaching in early childhood settings**. The content of the portfolios was then organized to support the related course goal. In the language arts methods course, the students had to develop a literacy portfolio to document their understanding of themselves as language users. They had to provide evidence in the following areas:

- earliest memories about learning to read and write
- reading for different purposes
- writing for different purposes
- using all stages of the writing process
- mastery of a word processing system
- oral interpretation of children's literature
- significant literacy role models in their own experience
- talking and listening for different purposes
- interpreting data in the above areas to
  - show awareness of the dimensions of their own literacy
  - describe their roles as a teacher of literacy
  - project their own further literacy goals

In the early childhood course, the students had to document that they:

- engaged in professional self-awareness and evaluation
- interacted with peers about professional issues

- articulated a personal philosophy of early childhood education
- projected goals and strategies to further empower themselves as early childhood professionals

Each of us developed our own portfolios along side of our students. Students were thus able to observe us as participant learners rather than as expert purveyors of information and assigners of tasks. This established a positive climate for the duration of the semester. Further, use of portfolios modeled for students the integration of assessment and instruction, a major concept that underpins the use of performance assessment. As students were documenting their learning in their portfolios, they were providing us with information that we could use to inform our instruction.

For example, in the language arts methods course, each of us shared our own **literacy autobiography** on the first day of class as we were explaining that the course requirements included the development of a literacy portfolio by each student. Later, a draft of each student's autobiography was read to us. We noted possibilities for library research topics suggested by the content of the draft. Then the writer shared the draft with a small group of peers to elicit further possibilities for research. By this point, the writer had several research topics to choose from as he/she initiated the research activity. The autobiography thus served two purposes: to document memories of early literacy experiences and to aid in the selection of a topic for a brief library research project.

Steps in Preparing Portfolios

At the first meeting of each of our respective classes, we showed our students the beginnings of our own portfolio. We showed artifacts that we had collected, told a little about each artifact and why it had been chosen, and read first drafts of reflections we had written about selected artifacts. We found that this modeling was effective and even critical. Students were immediately enthused about the prospect of collecting personal artifacts and later commented many times about the power of seeing the professor's personal involvement in the same process that they were to use.

To begin preparing their portfolios, we engaged our students in the steps of collecting, selecting, reflecting, and projecting (Hansen, 1992). First, students collected artifacts that they felt represented significant elements of their own literacy development or of their roles as early childhood educators, depending on the course in which they were enrolled. They brought these artifacts to class and shared them with one or two peers, discussing them informally. Artifacts included items such as books they had read as children, diplomas, papers written for previous classes, and so forth. Next, the students selected some of these artifacts to reflect on in writing. At another class meeting, students brought drafts of their written reflections and the accompanying artifacts to share with small groups of peers in a writing conference format. This cycle of collecting, selecting, and reflecting continued over the course of the semester, as students discarded some of their earlier choices of artifacts and added new ones.

At the same time, these processes were stimulating thought about areas of personal and professional development that many students had never considered before. They began to identify specific areas of strength and weakness in their own development and current functioning. Some students identified areas of personal strength such as a deep love of reading and keeping personal journals. They identified areas of weakness such as reading only magazines or romance novels, reading only when it was assigned, never choosing to write, and lack of ability to write organized essays. This deep reflection led naturally to the last step that we used, projecting. Students projected personal and professional goals for themselves based on the strengths, needs, and priorities that they had identified.

Management Strategies

In the language arts classes, peer conference sessions regarding written reflections on artifacts were organized around a simple strategy for providing useful feedback, informally dubbed "PQSC." In this strategy, students read or listened to a peer's written reflections, then provided specific feedback. First, they provided **positive** comments (P) to the writer about some aspect of the content of the draft. Next, the students asked genuine **questions** (Q) about some part of the content that they wanted to know more about land provided a specific **suggestion** (S) that the writer might want to consider as he or she revised the draft. Finally, the writer made a **commitment** (C) to any revisions he or she wanted to try. The students provided this structured feedback to each other on 3 x 5 inch note cards at the time of the peer conferences. The text of one student's "PQSC" note card from a conference on the student's literacy autobiography is reproduced below.

For Emily, by Tonya:

P: I like the part you added about learning to write in cursive the summer before third grade.
Q: Why did you move?
S: You could elaborate more about your family.
C: (by Emily) I will elaborate more on my family and I will expand in more detail on my literacy knowledge.

Additionally, the instructor used the same format to provide written feedback to students about their portfolios as they were in the process at interim evaluation periods.

As in the language arts classes, the instructor in the early childhood class requested that the students submit their portfolios for interim reviews. Peer conferences were scheduled during class. Again, comments from peers and the instructor proved to be useful to the students when selecting artifacts and writing accompanying reflections. The instructor provided written comments to each student concerning their written reflections. The interim reviews also provided the instructor with insights about the students' understanding of the assignment.

An even more comprehensive process evaluation was provided one time during the semester. Students brought their uncompleted portfolios to class and conferenced with small groups of peers, this time using forms that we designed that specifically listed the areas to be addressed in the respective courses. The peers provided feedback to the owners of the portfolios about which areas seemed to be well-covered and which areas still needed work. They also brainstormed possible artifacts that might address those areas that still needed to be covered or explained. Likewise, we used the same forms to provide written feedback to the students about the completeness of their portfolios at that point.

Students were also encouraged to share their portfolios with other significant people in their lives, and to include a special area in the portfolio for written comments from anyone who read the portfolio. Some students shared their portfolios with their parents, spouses, and children, and included comments from the readers in this section.

From the beginning of the semester, students created tables of contents for their portfolios, listing artifacts and the date that each artifact was added to the portfolio. If they removed any artifacts, the date of removal was added to the table of contents, so there was a permanent record of artifacts that had once been part of the portfolio but were no longer included. At the time the portfolios were turned in for evaluation at then end of the term, a specific presentational format was required. First, the table of contents were included, which directed the reader to the artifacts and reflections. At this point, each artifact selected for inclusion was accompanied by a written reflection which had been thoroughly revised and edited. The artifacts and reflections were followed by written interim feedback from peers, instructors, and others. The last section of the portfolios included final reflective summations that comprised what students had learned from their involvement in the portfolio process. They also projected personal goals related to their professional development as future teachers.

Evaluation of Portfolios

In the language arts classes, the instructor selected areas to address in the portfolio. As a result, a criteria sheet was constructed which reflected the required content of the literary portfolio. Product evaluation included a checklist of required areas along with qualitative ratings. Thus, the criteria sheet for the literary portfolio served as a road map for both the instructors and the students during the construction and the evaluation of the portfolio. In addition, the criteria sheet included sections addressing process evaluation as well as product evaluation. Process evaluation afforded the students with scheduled opportunities to share their portfolios in progress with their peers and the instructor before final submission. Moreover, the interim reviews provided the students with varied feedback to consider when making final revisions.

Student choice was evident in both classes with regard to the selection of artifacts for inclusion in the portfolios. In addition to the self-selection of artifacts, students in the early childhood class collaborated with the instructor to determine the criteria for final evaluation of their portfolios. In both classes, student decision making was an important component in the portfolio process.

Since multiple forms of assessment were used in each course, the portfolios represented only a portion of the final grade. In the language arts classes, course evaluation included the portfolio, a children's literature response journal and project, a library research paper, and exams. Assessment and evaluation criteria for the early childhood class included the portfolio and such items as a response journal, a home learning activity, a thematic unit of study, and exams.

At the end of the semester, we met to reflect upon our inclusion of portfolios in teacher education courses. This proved to be the final step in the instructor assessment and evaluation cycle for the classes. We agreed, first of all, that we got to know our students better including, information about their families, their aspirations, why they wanted to become teachers; and the power of the home-school connection. We further concluded that this assignment was a way to simultaneously integrate content, process, and strategy.

We still have questions concerning assessment and evaluation. However, our overall experience has confirmed our belief that assessment informs instruction. The interim reviews served as benchmarks when planning subsequent class sessions. Scheduled peer conferences provided the students with numerous opportunities to interact with ideas and materials. Evaluation criteria whether set forth by the course instructor or selected through collaboration with the student articulated expectations for the portfolio assignment. By participating in the portfolio process, the students were given opportunities to engage in reflective thinking, interact with peers, make informed decisions, and set personal learning goals. Hence the use of portfolios in our teacher education courses promoted the furtherance of the empowered person and professional as exemplified in the following reactions from students about the portfolio experience:

> "It made me take a look at myself."
> "This was my favorite assignment because it was the most personal."
> "You can add to it. As you grow. Add more artifacts and reflections."
> "Good exercise, helpful to find out who you are and putting goals on paper was useful."

Recommended Guidelines

The following guidelines for initiating portfolios in teacher education courses evolved from our experience and research inquiry:

- Select the purpose(s) for the portfolio based on the course goals
- Identify categories of evidence to include in the portfolio
- Model by participating in the construction of your own portfolio and share it with your students
- Engage students in the processes of collecting, selecting, reflecting, and projecting (Hansen, 1992)
- Ensure that each artifact is accompanied by a reflection
- Determine strategies for organization of the data in the portfolio (e.g., table of contents, response section, etc.)
- Collaborate with students when developing criteria for assessment and evaluation of the portfolio
- Use the portfolio process to inform instruction

REFERENCES

Hansen, J. (1992, August). *Literacy portfolios.* Paper presented at the meeting of the Gulf Coast conference on the Teaching of Writing, Point Clear, AL.

Paulsen, F.L., Paulsen, P.R., & Meyer, C.A. (1991). What makes a portfolio a portfolio? *Educational Leadership, 48*(5), 60-63.

Valencia, S. (1990). A portfolio approach to classroom reading assessment: The whys, whats and hows. *The Reading Teacher, 43*(4), 338-340.

Wolf, K. (1991). The school teacher's portfolio: Issues in design, implementation, and evaluation. *Phi Delta Kappan, 73*(2), 129-136.

Modeling Alternative Assessment in Teacher Education Classrooms

by

Mary Elizabeth D'Zamko and Lynn Raiser
University of North Florida

Modeling Alternative Assessment inTeacher Education Classrooms

Teachers for the today's schools need a new kind of teacher education to prepare them to meet the need of increasingly diverse students. Many schools have adopted cooperative learning and alternative assessment strategies to help students develop their creative problem-solving abilities and group process skills valued by today's business and industry (Johnson & Johnson, 1980; Slavin, 1983). As teacher educators, the authors have observed that many teachers are uncomfortable moving from teacher directed whole group instruction and traditional testing to cooperative learning and alternative assessment. Some teachers are resistant to changing their approach and do not give themselves and their students adequate time to learn new teaching and assessment strategies.

The authors incorporated cooperative learning methods in their university teacher education classes fully expecting that by using the methods, preservice and inservice teachers would find the strategies successful and satisfying , leading them to committed use with their own students. Instead we found that adult students continued to be unenthusiastic about group work and few seemed committed to adequate exploration of cooperative learning with their own students.

An analysis of our experience led us to the realization that what was missing was a strong theoretically-based approach to teaching that took us from being traditional "up front" teachers who used cooperative learning to spice up class activities to teachers willing to share the power with our students. Until we shared the power and became true teaching and learning collaborators with our students, there was no excitement about cooperative learning in our university classrooms.

The Foxfire Approach

In our search for a collaborative approach to teaching and learning we discovered the Foxfire strategy developed over the last 25 years by Eliot Wigginton at Rabun Gap High School in the north Georgia foothills of the Appalachian Mountains. The Foxfire approach is based on the constructivism philosophy which views people as capable learners constantly creating new meanings in ways that fit their own experience, interests, developmental levels and prior learning. Learning is ot a simple process of absorbing new information that exists externally. Rather, learning is "the natural, continuous construction and reconstruction of new, richer, more complex and connected meanings by the learner" (Poplin & Stone, 1992, p. 161).

Instruction based upon the constructionist view focuses on student self-management and self-regulation with learners actively involved in the learning process. It embodies the principles of Dewey (1938) emphasizing experience, experiment, freedom, democracy and purposeful learning. Dewey purported that education is a social process; a process of living, not mere preparation for future living.

The Foxfire approach also accommodates the different talents and abilities of students (Gardner, 1983). Students are encouraged to use their talents in presentations and other class activities. The approach incorporates best practices in special education, allowing students to learn and demonstrate mastery in ways that reflect their unique abilities, interests and learning styles (Ensminger, (1992).

During the summer of 1992 we joined thirty K-12 teachers in Rabun Gap in a Level One Foxfire course to learn how to be collaborative teachers. We returned to our classes at the University of North Florida with the Eleven Foxfire Core Practices (*Foxfire Approach*, 1991) and began collaborations with our students in which they made choices about how they would learn the course objectives and how they would be evaluated.

Foxfire Course Organization

Students are given the course objectives and some assignments that are nonnegotiable "givens." Planning and scheduling the activities to meet the objectives is done collaboratively with the students. Among the strategies planned collaboratively are: cooperative learning teams, learning partners (study buddies), simulations, role plays, discussions of films and other media, group and individual presentations, learning games, computer exercises, and guest speaker discussions. The instructor, as course facilitator, provides procedural and content information and serves as a resource during their work. The instructor is an *ex officio* member of each learning team.

Methods of Assessment

Assessment flows naturally out of course activities. Teaching, learning and assessing are continuous processes as students actively construct their own meaning in collaboration with the instructor and other students. Students develop collaborative projects with fellow students and create a Process Folio to document their personal learning.

Responsive constructionist evaluation is used to assess progress in the course. Guba and Lincoln (1989) refer to this alternative evaluation process as fourth generation evaluation.

First generation evaluation uses tests to measure whether students have mastered the content of a course. Tests are also used to measure aptitudes for classifying students for various purposes (programs for gifted or mentally handicapped students, college entry, etc.). Whereas first generation evaluation targets students, second generation evaluation assesses curricula by describing whether or not students are learning what the instructor is teaching. In order to do this, desired learning outcomes (objectives) are established and evaluators describe strengths and weaknesses in curricula without judgment. Third generation evaluation adds judgment to second generation evaluation by establishing standards against which program objectives can be measured.

Fourth generation evaluation, proposed by Guba and Lincoln (1989), engages all the stakeholders in the evaluation process. Students openly negotiate from equal positions of power with the instructor. Students discuss the objective and assignments and compile a personal Process Folio. "A portfolio is a purposeful collection of student work that exhibits the student's efforts, progress, and achievements in one or more areas. The collection must includes student participation in selecting contents, criteria for selection, the criteria for judging merit, and evidence of student self-reflection" (Paulson, Paulson, & Meyers, 1991, p. 60).

Students in the course determine the contents of the Process Folio with the exception of some exercises that are required by the instructor. Among these requirements is a collection of individual reflections which focus on the student's reactions to topics from the text, readings, discussions in group work in class. Clift, Houston, and Pugach (1990) suggest that the teacher educators should be concerned with developing reflective practitioners.

Howard Gardner (Brandt, 1988) describes the contents of a portfolio as an accumulation of various efforts "including drafts, notes, false starts, things they like and don't like" (p. 33). The portfolio becomes a collection of what's been done and what's been learned that both teacher and student can examine for progress.

The purpose of the Process Folio is to show the interrelationships and realities that the student is constructing as he/she progresses through the course. The Process Folio also serves as a vehicle for content and assessment dialogue between the student and the instructor, leading to the ultimate purpose of evaluation which is to enable students to evaluate themselves Costa (1989).

The Foxfire Approach used in the courses is predicated on Eleven Core Practices, three of which directly relate to assessment:

Core Practice 3: "The academic integrity of the work must be absolutely clear" *Foxfire Approach,* p. 3).

Core Practice 10: "Reflection—some conscious, thoughtful time to stand apart from the work itself—is an essential activity that must take place at key points throughout the work" (p.4).

Core Practice 11: "The work must include unstintingly honest, ongoing evaluation for skills and content, and changes in student attitudes" (p. 4).

The overriding question is, "In what ways will you prove to me at the end of this class that you have mastered the objectives it has been designed to serve?" *Foxfire Approach*, p. 4). Students set personal goals, monitor their own progress, and help others in the class progress in their learning. The classroom climate is supportive and non-competitive. There is no concern for grade distribution, curving grades or ranking students. Every student is a person of value to all members of the class.

Components of the Courses

1. Students set personal goals for meeting course objectives
2. Students write dialogues with the textbook and other written material, commenting on insights as they read.
   EXAMPLES:
   Comment on tables
   Xerox diagrams and map concepts
   Prepare questions over a chapter
   Develop learning games for the class

3. Students write reflections, personal responses to all aspects of course.
    EXAMPLES:
    Responses to teaching method used in class
    Responses to effectiveness of group processes, including diagnoses of group member interaction problems
    Responses to articles and other personal, non-assigned reading
    Responses to relevant television, movies, and other media
    Responses to cartoons, newspaper articles
    Responses to their own classroom teaching experiences
    Individual research in areas of interest
4. Students participate in individual and group projects
    EXAMPLES:
    Group review and sharing of reading tests
    Reviews of curriculum materials
    Demonstrations of computer-assisted instruction
    Group creation and presentations of whole language units
    Review of learner characteristics and implications for teaching
5. Students have free choice of any other activities that will answer the question: "What am I learning as a result of this course?"
    EXAMPLES:
    Affecting positive changes in attitudes towards schools
    Personal plan to defuse Teacher's Lounge Syndrome of negativity
    Assessing personal growth in working with others
    Dealing with stress
    Optional quizzes
    Self-evaluation
    Establish evaluation criteria for instructor to use
    Demonstrations such as sharing videos of one's class
    Peer evaluations

We are continually learning, doing better with some aspects of the Foxfire approach than others, as we move from directing students to learn what we teach to guiding students to construct their own meaning from course experiences. Although the collaborative learning seems to run smoothly, assessment continues to be a challenge. At this point in our limited practice, we are searching for standards to assess Process Folio work. We have asked students to write proposals to guide our assessments of their work. We tend to evaluate holistically and, although we are committed to individualizing assessment, we find it difficult not to compare students with each other. We can tell early in the courses who the superior students are. We strive to give more structure and guidance to the others so that they can achieve the high standard set naturally by the few who do outstanding work from the beginning.

With experience, we hope to tighten criteria and develop methods of quantifying Process Folio work, group projects and other non-traditional demonstrations of mastery of course objectives. As we refine our alternative assessment processes, we will continue to use the Foxfire approach which has liberated not only our students to construct their own meaning, but has freed us to explore teaching and learning in richly satisfying collaborations with our students who constantly enlighten and surprise us with their wisdom.

REFERENCES

Brandt, R. (1988). On assessment in the arts: A conversation with Howard Gardner. *Educational Leadership,* December, 1987/January, 1988, 30-34.

Clift, R.T. Houston, W.R., & Pugach, M.C. (Eds.) (1990). *Encouraging reflective practice in education: An analysis of issues and programs,* New York: Teachers College Press.

Costa, A.L. (1989). Reassessing assessment. *Educational Leadership,* 46(2), 2-3.

Dewey, J. (1938). *Experience and education.* New York: Collier Macmillan.

Ensminger, G. (1992). The Foxfire pedagogy: A confluence of best practices for special education. *Focus on Exceptional Children,* 24(7).

Gardner, H. (1983). *Frames of mind: The theory of multiple intelligences.* New York: Basic Books.

Guba, E.G., & Lincoln, Y.S. (1989). *Fourth generation evaluation.* Newbury Park, CA: SAGE.

Johnson, R.T., & Johnson, D.W. (1983). Effects of cooperative, competitive, and individualistic learning experience on social development. *Exceptional Children,* 49, 323-329.

Paulson, F.L., Paulson, P.R., & Meyer, C.A. (1991). What makes a portfolio a portfolio? *Educational Leadership,* 48(5), 60-63.

Poplin, M.S., & Stone, S. (1992). Paradigm shifts in instructional strategies: From reductionism to holistic/ constructivism. In W. Stainback and S. Stainback (Eds.) *Controversial issues confronting special education,* Boston: Allyn & Bacon.

Slavin, R.E. (1980). Cooperative learning. *Review of Educational Research,* 15, 315-342.

The Foxfire approach: Perspectives and core practices. (1991). *Hands on: A Journal for Teachers,* 41.

Part 3

An Analysis of Curriculum Domains: Implications for Assessment, Program Development and School Improvement

Linda S. Behar, Ph.D.
Assistant Professor
College of Education
Department of Educational Leadership
2412 Norman Hall
University of Florida
Gainesville, FL 32611
(904) 392-2391 Ext. 269
(904) 392-0038 FAX

An Analysis of Curriculum Domains: Implications for Assessment, Program Development and School Improvement

Introduction

Defining professional standards and establishing accountability measures to ensure that graduates have acquired the skills and behaviors necessary for successful practice in the real world of curriculum work is integral to schools' courses in implementing improvement plans. Departments and colleges of education can provide leadership and assist schools in their improvement initiatives by establishing criteria for the emerging curriculum specialist. This study reports the findings related to establishing a knowledge base of important curriculum domains and related curriculum practicers that were identified in a content analysis of curriculum textbooks[1] published between 1970-1990. The identification of this knowledge base was supported by quantifiable procedures.

Knowledge Bases

The field of education, particularly curriculum is undergoing tremendous transitions in an effort to identify an operational knowledge base. Knowledge bases that are pertinent to curriculum studies may be conceptualized in terms of classical topical categories, research domains, and paradigms of teacher education (Gideonse, 1989).

Knowledge bases include different ways of knowing that are important for professional educators and necessary for practice (Gudmundsdottir, 1991). A knowledge base can be developed from source documents such as a body of bibliographic sources (textbooks) that summarize key concepts and principles from research, theory, and practice. Selected from source documents, knowledge bases provide a theoretical framework that is comprised of essential knowledge, current research findings and sound practice to provide a structure for making informed decisions. Central to the formulation of a knowledge base are the development of beliefs about purposes of schools, roles of teachers, educational philosophies, theories and research, social perspectives, educational practices, research on teaching and contemporary societal concerns.

The knowledge base of curriculum can be defined in terms of **domains** of the field, that is broad concepts, as well as practices that help define the field. The kind of work that curriculum specialists do has been described as activities, behaviors, or roles. The referent, **curriculum practices**, is used herewith to represent the behaviors and activities that help define what curriculum workers do in the real world of planning, implementing, or evaluating curriculum. I have postulated that the domains and practices are interrelated and that curriculum practices, the behaviorist or specific aspect of the field can be used to define and operationalize the broader, more abstract (domains) parts of the field.

Domains of Curriculum

Domains of curriculum are ways of structuring the "knowledge base" of a field of study or a professional discipline. They can be viewed in philosophical and/or operational terms. Domains are important content areas or classical topical categories within a discipline that researchers and text authors examine in an attempt to further the field of knowledge. Domains represent ways of structuring the knowledge base of a field of study and establishing modes of inquiry. They suggest broad conceptualizations of curriculum that yield specific curriculum activities. By delineating the curriculum domains, we can establish the means-ends process and assumptions to decision- making in curriculum. The domains of curriculum were systematically selected from bibliographic sources that served to undergird their inclusion and promulgate essential knowledge in the areas of theory, research and practice. A content analysis of curriculum textbooks published between 1970 and 1990 revealed eleven domains of curriculum: (1) curriculum philosophy, (2) curriculum theory, (3) curriculum research, (4) curriculum history, (5) curriculum change, (6) curriculum development, (7) curriculum design, (8) curriculum implementation, (9) curriculum evaluation, (10) curriculum policy, (11) curriculum as a field of study.

---

[1] A textbook is a book used for the study of a particular subject. It is a book designed to explain basic information of a field, including theory, research and practice. For the purposes of this study, the reference to textbooks is limited to books in the field of curriculum.

Curriculum philosophy is defined as a set of values, beliefs, and/or a particular orientation that determines an individual's broad view of a subject. It guides students, teachers, and schools in both teaching and learning. Inquiry into educational philosophy suggests a general view of students, society, as well as curriculum. Educational philosophy leads to a determination of educational theory, educational aims, and curriculum development and design.

Pertinent to the aims of curriculum philosophy are determining how conceptions of human nature, society and values influence the views of education. This domain examines the quality of education, the meaning of equity in education and explores the standards determined by personal, social, and national concerns that should be met by schools (Ornstein & Levine, 1989). Curriculum philosophy helps educators answer value-laden questions and make decisions among many choices.

The literature identifies five educational philosophies: 1) perennialism; 2) essentialism; 3) progressivism; (4) reconstructionism; and (5) existentialism (Doll, 1989; Ornstein & Hunkins, 1993).

Curriculum theory is defined as a set of related statements that give meaning to a school's curriculum by highlighting the relationships among important elements and by directing its development, use, and evaluation (Beauchamp, 1961).

Curriculum theory uses techniques of science and logic to identify fairly stringent rules that present a systematic view of phenomena. It is an activity that involves theorizing and reflecting which can also be interpreted as a process of clarifying meaning and the use of language (Schubert, 1986). McNeil (1990) divides the curricular theorists into two camps, the soft and the hard curriculum theorists. The soft curricularists are concerned with understanding and revealing the political and moral aspects of curriculum making. Soft curricularists do not study change in behaviors or decision making in the classroom. They are concerned with concepts of temporality, transcendence, consciousness, and politics and their relationship to the process of education.

The hard curricularists assume that curriculum development occurs in response to an idea or vision of what ought to be taught. A series of logical choices or scientific justification determines the curriculum design. Empirical confirmation is the basis for justification.

Curriculum research is an activity used to: 1) advance conceptualizations and understanding of the field; 2) create new visions of what and how to teach; 3) influence curriculum policy; 4) question normative premises about curriculum; and 5) improve programs for learning (McNeil, 1990). Considered a mode of systematic inquiry for the purpose of solving a particular curriculum problem, curriculum research analyzes the steps to be taken in solving a given problem, tries one or more actions in line with that analysis, and then observes whether actions brought the results that were predicted or anticipated in the analysis (Doll, 1989).

Curriculum history is the process of describing, analyzing, and interpreting past curriculum thought and practice. Like history, it is a chronicle record of past events that may be represented by a narrative and/or an analysis of past events. By analyzing the past and the origins of curriculum, educators can better understand the present. A study of curriculum history can reveal insight and approaches to problems that relate to similar present day issues. An investigation of the forces that inhibited or promoted particular curriculum innovation, decisions, and action in the past can help educators analyze present conditions and plan future course of action (Schubert, 1986).

Curriculum change is an activity geared towards curriculum improvement. Curriculum developers are challenged with getting curriculum adopted at national, state, and local levels. Their plans must be accepted by textbook committees, curriculum commissions, boards of education, and others so that curriculum can be made available to teachers (Saylor, Alexander, & Lewis, 1981). Insuring that curriculum changes are properly implemented is another task. Some teachers might not be able to enact curriculum changes developed by others. Implementing curriculum change should take into account the special knowledge and suggestions of those directly responsible for enacting the curriculum innovations (McNeil, 1990). For curriculum change to begin and endure, strategies for achieving cultural or institutional change are more significant than strategies for achieving technological change (Doll, 1989).

Curriculum development is the process of deciding what to teach and learn, along with the considerations needed to make such decisions (Schubert, 1986). Integral to this effort is the identification of tasks, steps, roles, expectations, resources, time and space, and the ordering of these elements into a system for carrying out the specified design to create a curriculum plan or document (Kimpston & Rogers, 1986). Curriculum development is an activity that determines how curriculum construction will proceed. The process addresses the questions of who will be involved in curriculum construction and what procedures will be used in this process.

Curriculum design sometimes called curriculum organization is the arrangement of curriculum into substantiative entities. Generally, it consists of four components: (1) aims, goals, or objectives, (2) content, (3) learning experiences and (4) evaluation approaches. Sources for curriculum design are the learner, science, society, knowledge, and in some cases the external or divine. Specific design dimensions include scope, articulation, balance, integration, sequence, continuity, and interrelatedness (Ornstein & Hunkins, 1993).

Curriculum design is a way of organizing curriculum ideas so they function in the real world of classrooms and schools. It might also be considered a carefully conceived plan that takes into account what its creators want done, what subject matter will be used, what instructional strategies will be used, and how the designer will determine the success or feasibility of the design. Diagnosis of need, organization and selection of both subject matter and learning experiences are usually related tasks of curriculum design (Doll, 1989).

Curriculum implementation refers to the planning for and actual use of a curriculum in practice and concerns the process of putting into effect the curriculum that was produced by construction and development (Kimpston & Rogers, 1986). Curriculum implementation by definition offers evaluative feedback to those in charge of the construction and developmental processes (Giroux, Penna & Pinar, 1981).

Curriculum implementation can be defined as a system of engineering that takes design specifications through various channels to the teacher and the classroom. It can be an interpretation of how well teachers carry out instruction in a school district. Curriculum implementation can refer to the development of learning experiences based on knowledge derived from the continuous interactions with learners (Schubert, 1986).

Curriculum evaluation is the process of answering questions of selection, adoption, support and the worth of educational materials and activities (Scriven, 1967). Integral to curriculum evaluation is an emphasis on improving the curriculum (Stufflebeam, 1971).

Tyler (1949) delineates the task associated with curriculum evaluation as: (1) determining the effectiveness of curriculum content; (2) measuring discrepancies between predetermined objectives and outcomes; (3) providing information about students' needs and capabilities; (4) guiding program development for individual students; (5) providing information about the success or effectiveness of curriculum content; (6) determining if objectives have been met and what changes took place as a result of the curriculum; (7) identifying strengths of curriculum content; (8) offering suggestions for modification, and (9) specifying curricular changes that need to be made with respect to content, instructional strategies, or methods that might lead to more effective curricular implementation.

Curriculum evaluation serves several purposes: 1) it provides a periodic assessment about the effectiveness of the curriculum indicating changes that will facilitate improvement; 2) it influences teaching and learning by offering data essential to guiding individual students; and 3) it can validate hypotheses upon which curriculum selection and implementation operates (Madaus and Stufflebeam, 1989).

Curriculum evaluation is a continuous process that ascertains whether the planning, monitoring, and reporting of curricular activities regarding persons, procedures, and objects involved in actual situations have been achieved (Giroux, Penna & Pinar, 1981).

Curriculum policy is usually a written statement of what should be taught and serves as a guide to curriculum development. It establishes ground rules, limits, and criteria that circumscribe the curricula of educational institutions within a given jurisdiction. Curriculum policy must be determined by a democratic process whereby the wishes of all concerned parties are considered before legalization (Saylor, Alexander, & Lewis 1981).

An authoritative allocation of competing values, curriculum policy addresses issues regarding graduation requirements, mandatory curriculum, and frameworks outlining the content for a field of knowledge (McNeil, 1990). Curriculum policy also addresses the question of what groups should influence the curriculum and to what extent. A mandate is decision to promote one goal over another is an example of curriculum policy (McNeil, 1990).

Curriculum as a field of study is the combination of curriculum, the curriculum system, and research and theory building activities (English, 1983). Curriculum is the substantive or content dimension of curricular planning, implementing, and evaluation.

Zais (1976) defines curriculum as a field of study as the range of subject matters with which it is concerned (the substantive structure) and the procedures of inquiry and practice that it follows (the syntactical structure). The curriculum field may be described as the subject matters that are treated in schools and the processes (for example, curriculum development, and curriculum change) with which specialists are concerned (Giroux, Penna & Pinar, 1981).

According to Ornstein (1987), curriculum as a field of study consists of its own foundations, domains of knowledge, research, theory, and principles.

Research on Curriculum Domains

Several experts have underscored the lack of agreement in defining the curriculum domains. Beauchamp (1961) was one of the first theorists to analyze curriculum in terms of domains, which he called "curriculum knowledge," into planning, implementing and evaluating. Foshay and Beilin (1969) used the term curriculum knowledge and divided curriculum into theory, design, and change. Rosales-Dordelly and Short (1985) established what they called the "conceptual framework" and identified eight "specialized knowledge" areas of the field: policy making, development and evaluation, change and enactment, decision making, field of study or an activity, forms of inquiry, languages for inquiry and questions directing activity. They concluded that the body of curriculum knowledge was incoherent, and fragmented. More recently, Ornstein and Hunkins 919930 concluded that the only agreed-upon and traditional domains were curriculum development and curriculum design—the technical aspects of curriculum construction.

Up to this point, all of the above constructs dealing with curriculum knowledge and curriculum domains lacked empirical support. These concepts or ideas were based solely on language and qualitative discussions. The fact that there is considerable disagreement about curriculum domains suggests that the field lacks an agreed upon theoretical base.

Curriculum Domains Quantified

For this study, as many as nine curriculum domains were identified and quantified with related practices or behaviors that curriculum specialists perform. The domains were: (1) curriculum philosophy, (2) curriculum theory, (3) curriculum research, (4) curriculum history, (5) curriculum development, (6) curriculum design, (7) curriculum evaluation[2], (8) curriculum policy, and (9) curriculum as a field of study.

Each domain was defined by four or more curriculum practices (items) by two elementary teacher groups, an urban sample in the North, the Illinois teacher group, (**ILTCHR**), N=48, and a rural sample in the South, the Florida teacher group (**FLTCHR**), N=37, and Professors of Curriculum (**PROFCURR**), N=51[3]. The curriculum practices were representative of the kinds of activities performed by curriculum specialists (including teachers, principals, coordinators, and directors of curriculum).

The domains and practices were quantified through formal reliability and validity procedures. In phase one of the study, a group of experts (N=5) independently categorized a list of 89 domain practices into one of eleven domains to help establish content validity. As a result of this categorization process, two curriculum domains (curriculum change and curriculum implementation) were omitted because an insufficient number of curriculum practices were categorized within the domains. Thirty-four domain practices were omitted as a result of this stage.

For the purposes of establishing reliability, phase two, the Illinois teacher group (N=48) and Professor of Curriculum group (N=51) each rated the importance of the curriculum practices on a five-point Likert scale from "very important" to "very unimportant" with the midpoint being "of some importance." In order for each curriculum practice to be included in instrument it had to exhibit an item discrimination score of .20 or higher within its respective system (or subscale) as per the teacher ratings and professor ratings. Six domain curriculum practices had item-total scores below .20. As a result, these items were eliminated at this stage because of insufficient alpha correlation coefficient scores. The curriculum practices and related domains eliminated as a result of Professors of Curriculum ratings were: (1) determines what changes took place as a result of the curriculum (domain: Curriculum evaluation), (2) influences control of the curriculum (domain: curriculum policy), (3) recommends what learning experiences to include (domain: curriculum policy), (4) develops curriculum guides (domain: curriculum development), (5) develops school grants (domain: curriculum development), and (6) addresses questions of who will be involved in curriculum construction (domain: curriculum development).

In phase three, the Florida teacher group (N=37) also rated the importance of the curriculum practices using the five-point Likert scale. As a result of these findings, 3 curriculum practices were deleted because they had alpha coefficients of less than .20. Curriculum practices and corresponding domains eliminated by the Florida teachers included: (1) determines ends of education (domain: curriculum philosophy), (2) determines what changes took place as a result of the curriculum (domain: curriculum evaluation)[4], and (3) communicates with local and state government agencies (domain: curriculum policy). The reliability data obtained in this third application represents a further refinement of the instrument. The remaining 47 curriculum practices within their respective 9 subscales (curriculum domains) are shown in Table 1.

---

[2] Curriculum evaluation as it pertains to a domain is a micro-level process that is concerned with the effectiveness of the curriculum in the classroom and focuses on a means-end assessment.

---

[3] The Professors of Curriculum are specialists who conduct research and are consultants to schools and education agencies. They are elected to membership by invitation because of their significant contributions and/or publications in the field of curriculum studies.

[4] This curriculum practice was also eliminated by the Professors of Curriculum

---

TABLE 1.

CORRECTED ITEM-TOTAL CORRELATIONS AND ALPHA COEFFICIENTS FOR THE IMPORTANCE OF CURRICULUM PRACTICES WITHIN THE CURRICULUM DOMAINS BY THE ILLINOIS TEACHER GROUP (**ILTCHR**), FLORIDA TEACHER GROUP (**FLTCHR**) AND THE PROFESSORS OF CURRICULUM (**PROFCURR**).

DOMAINS OF CURRICULUM

	CORRECTED ITEM-TOTAL CORRELATIONS		
	ILTCHR	FLTCHR	PROFCURR
	(N = 48)	(N = 37)	(N = 51)
I. CURRICULUM PHILOSOPHY	a=.7307*	a=.6435*	a=.8450*
Curriculum Practice			
1. Schools of thought including : perennialism, essentialism, progressivism, reconstructionism, and existentialism.	.2660	.3206	.7025
Determines the ends of education.	.3748	b	.4880
2. Determines and orientation to curriculum.	.4228	.2925	.6799
3. Suggests a view of society and students. in relationship to education.	.4873	.5286	.5323
4. States the purposes of education.	.6670	.4926	.6428
5. Elaborates on the theory of curriculum.	.6337	.3200	.7101
II. CURRICULUM THEORY	a=.8306	a=.6950	a=.6974
Curriculum Practice			
6. Creates statements that give meaning to a school curriculum.	.5470	.4916	.6467
7. Uses techniques of science and logic to present a systematic view of phenomena.	.6930	.5410	.4298
8. Deals with structuring knowledge.	.6202	.4442	.4969
9. Identifies how students learn.	.6509	.5722	.4237
10. Uses principles and rules to study curriculum.	.6393	.2364	.2630
III. CURRICULUM RESEARCH	a=.8468	a=.6920	a=.7340
Curriculum Practice			
11. Analyzes resisting and supporting forces.	.7320	.3308	.4059
12. Advances hypotheses and assumptions of the field.	.6502	.4699	.5783
13. Uses systematic inquiry for the purpose of solving a particular problem.	.7192	.5804	.4473
14. Analyzes steps to be taken in problem solving.	.5778	.4269	.5201
15. Focuses on research and/or inquiry of curriculum.	.5993	.4307	.5243

	CORRECTED ITEM-TOTAL CORRELATIONS		
	ILTCHR	FLTCHR	PROFCURR
	(N = 48)	(N = 37)	(N = 51)
IV. CURRICULUM HISTORY	a=.7884	a=.6473	a=.7580
Curriculum Practice			
16. Describes past curriculum thought and practices.	.6290	.4591	.4127
17. Interprets past curriculum practice.	.6500	.4598	.7323
18. Provides a chronology of important event in curriculum.	.5052	.3370	.5725
19. Examines forces that inhibit curriculum innovations.	.5932	.4639	.2322
V. CURRICULUM DEVELOPMENT	a=.8695	a=.7278	a=.6236
Curriculum Practice			
Develops curriculum guides.	.7951	.3678	b
Develops school grants.	.7046	.4610	b
20. Determines procedures necessary for a curriculum plan.	.7317	.3120	.1988
Addresses questions of who will be involved in curriculum construction.	.5622	.5834	b
21. Integrates content and learning experiences.	.5797	.6075	.4917
22. Decides on nature and organization of curriculum.	.6551	.4889	.6499
VI. CURRICULUM DESIGN	a=.9049	a=.8663	a=.8505
Curriculum Practice			
23. Attempts to define what subject matter will be used.	.5282	.4894	.6288
24. Guides program development for individual students.	.7200	.4639	.7463
25. Selects subject matter and learning experiences.	.7408	.7088	.6173
26. Establishes the primary focus of subject matter.	.8568	.8427	.7389
27. Permits curriculum ideas to function.	.6524	.6025	.4871
28. Integrates careful planning.	.7818	.6624	.7631
29. Indicates instructional strategies to be utilized.	.7830	.7116	.3492

	CORRECTED ITEM-TOTAL CORRELATIONS		
	ILTCHR	FLTCHR	PROFCURR
	(N = 48)	(N = 37)	(N = 51)
VII. CURRICULUM EVALUATION	a=.9332	a=.8738	a=.8483
Curriculum Practice			
Determines what changes took place as a result of the curriculum.	.2521	b	b
30. Provides information about the effectiveness of the curriculum.	.5642	.3484	.3264
31. Determines whether actions yielded predicted results.	.7197	.4447	.4984
32. Determines if objectives have been met.	.8437	.6849	.4540
33. Offers suggestions for curriculum modification.	.7489	.6115	.2716
34. Measures discrepancies between predetermined objectives and outcomes.	.7268	.4538	.2727
35. Judges worth of instructional methods and materials.	.7419	.3860	.4624
36. Determines desired outcomes of instruction.	.7938	.4875	.6907
37. Improves curriculum programs.	.7506	.5714	.6040
38. Determines effectiveness of curriculum content.	.8234	.6981	.6923
39. Ascertains whether outcomes are the result of the curriculum.	.7085	.5979	.7627
40. Determines criteria to measure success of curriculum plan.	.7436	.8008	.6328
41. Identifies the strengths of curriculum content.	.7241	.7298	.5908
VIII. CURRICULUM POLICY	a=.7964	a=.6270	a=.7350
Curriculum Practice			
Influences control of the curriculum.	.5965	.3817	b
Recommends what learning experience to include.	.6605	.2252	b
42. Mandates school goals.	.7015	.4958	.5309
43. States what ought to be taught.	.5781	.5127	.6497
Communicates with local and state government agencies.	.3763	b	.4942
IX. CURRICULUM AS A FIELD OF STUDY	a=.8697	a=.7786	a=.7092
Curriculum Practice			
44. Promotes curriculum planning and implementation.	.7966	.6284	.2080
45. Organizes patterns and structures of curriculum.	.7637	.5774	.4157
46. Attempts to integrate theory and practice.	.6423	.5869	.6225
47. Analyzes structures of curriculum.	.6999	.5430	.4805

Notes: *a= The alpha correlation coefficient for each domain by teachers and professors, that is, how the curriculum practices correlated within their respective domains.
b= Denotes that item was eliminated because it had an item discrimination score of less than .20

TABLE 2.

NUMBER OF CURRICULUM PRACTICES WITHIN DESIGNATED ALPHA CORRELATION COEFFICIENT RANGES BY ILLINOIS TEACHERS **(ILTCHR)**, FLORIDA TEACHERS **(FLTCHR)** AND THE PROFESSORS OF CURRICULUM **(PROFCURR)**.

	ALPHA CORRELATION COEFFICIENT RANGE		
SAMPLE GROUP		.50 OR HIGHER	.20 - .49
ILTCHR N=48	49	6	
FLTCHR* N=37	25	27	
PROFCURR** N=51	28	21	

Notes: *= Excludes 3 curriculum practices that had item-total correlations of less than .20
**= Excludes 6 curriculum practices that had item-total correlations of less than .20

Table 1 shows the alpha coefficients for the importance of curriculum practices within the nine domains of curriculum.

Ratings by the Illinois Teacher Sample

Ratings for each of the subscales by the Illinois teacher sample demonstrated high levels of homogeneity. All nine of the domains evidenced alpha coefficients above of .70 or higher: curriculum evaluation (a=.93), curriculum design (a=.90), curriculum as a field of study (a=.85), curriculum theory (a=.83), curriculum policy (a=.80), curriculum history (a=.78), and curriculum philosophy (a=.73).

Overall, forty-nine curriculum practices had alpha coefficients of .50 or higher and six had ranges of .20 to .49. The Illinois teachers evidenced considerable agreement in their ratings of the curriculum practices. Table 2 indicates the Illinois teachers' tended to rate the curriculum practices higher than both the Florida teachers and the Professors of Curriculum. Perhaps this is a reflection of the type of curriculum work they are engaged in, or an indication that they work with colleagues who share similar values concerning the importance of various curriculum practices.

Ratings by the Florida Teacher Sample

High levels of homogeneity for the curriculum domains were demonstrated by the Florida teacher group. Ratings by the Florida teachers revealed that 6 of the curriculum domains had alpha coefficients of .70 or higher: curriculum evaluation (a=.87), curriculum design (a=.87), curriculum a a field of study (a=.70) and curriculum theory (a=.70). The remaining three domains had alpha correlations of..60 or higher: curriculum history (a=.65), curriculum philosophy (a=.64), and curriculum policy (a=.63). It should be noted that only 2 practices remained in curriculum policy. This under representation represents a limitation of the instrument and the overall contribution of this domain should be considered somewhat questionable.

Ratings for the domains by Florida teachers were slightly lower than ratings by Illinois teachers and the Professors of Curriculum, although each of the subscales evidenced high levels of homogeneity. Curriculum practices with alpha correlations scores in the range of .50 or higher totaled 25. Twenty-seven (27) curriculum practices had alpha coefficients of .20 to .49 (Refer to Table 2). Once again it should be noted that three curriculum practices which had alpha correlation coefficients of less than .20 were deleted from the instrument. The Illinois teachers' ratings of the curriculum practices were slightly higher than those by the Florida teachers. The differences in ratings between the Florida and Illinois teachers might be a reflection of differences in teacher preparation training, pedagogical values, and institutional factors that influenced what teachers considered important practices.

Ratings by the Professors of Curriculum

Regarding ratings by the Professors of Curriculum, all of the curriculum domains, except curriculum development (a=.62) evidenced substantially high levels of homogeneity. The remaining eight domains evidenced alpha coefficients of .70 or higher: curriculum design (a=.85), curriculum philosophy (a=.84), curriculum evaluation (a=.84), curriculum history (a=.76), curriculum research (a=.73), curriculum policy (a=.73), curriculum as a field of study (a=.71), and curriculum theory (a=.70).

Overall, excluding the six curriculum practices which had item discrimination scores of less than .20, twenty-eight curriculum practices had alpha coefficients of .50 or higher and twenty-one had ranges of .20 to .49 (See Table 2). An overview of the ratings of curriculum practices by the Professors of Curriculum and the Florida teachers demonstrated that there was a comparable number of items in the ranges of .50 and higher (28 and 25, respectively) and .20 to .49 (21 and 27, respectively). The fact that the Professors of Curriculum are not engaged in curriculum work on an everyday basis or work with colleagues who participate in similar roles and behaviors might account for the ratings they assigned to the curriculum practices.

TABLE 3.

THE IMPORTANCE OF CURRICULUM DOMAINS AS INDICATED BY ALPHA CORRELATION COEFFICIENT SCORES BY THE ILLINOIS TEACHERS (**ILTCHR**), FLORIDA TEACHERS (**FLTCHR**) AND THE PROFESSORS OF CURRICULUM (**PROFCURR**) IN ORDER OF RANK.

CURRICULUM DOMAIN	ILTCHR RANK N=48	ALPHA*	FLTCHR RANK N=37	ALPHA	PROFCURR RANK N=51	ALPHA
PHILOSOPHY	9	.7307	8	.6435	3	.8450
THEORY	6	.8306	5	.6950	8	.6974
RESEARCH	5	.8468	6	.6920	6	.7340
HISTORY	8	.7884	7	.6473	4	.7580
DEVELOPMENT	4	.8695	4	.7278	9	.6236
DESIGN	2	.9049	2	.8663	1	.8505
EVALUATION	1	.9332	1	.8738	2	.8483
POLICY	7	.7964	9	.6270	5	.7350
FIELD OF STUDY	3	.8697	3	.7786	7	.7092

Notes: Alpha*= Denotes alpha correlation coefficient score

Rank Order Importance of the Curriculum Domains

Table 3 shows that the Illinois and Florida teachers ranked curriculum evaluation, curriculum design, curriculum as a field of study and curriculum development in ranks 1 through 4 , respectively. The Illinois teachers rated curriculum research and curriculum theory in ranks 5 and 6; the Florida teachers rated the same domains in ranks 6 and 5. Illinois teachers ranked curriculum policy, curriculum history and curriculum philosophy in ranks 7, 8 and 9, respectively. The Florida teachers rated these three domains in ranks 9, 7, and 8, respectively.

The Professors of Curriculum ranked curriculum design and curriculum evaluation as 1 and 2. The Illinois and Florida teacher groups rated the respective domains as 2 and 1. Only one other similarity was noted between the Professors of Curriculum and the Illinois/Florida teacher groups rank order ratings of the domains. The professors rated curriculum research as rank order 6, concurring with the Florida teachers; the Illinois teachers rated this domain as rank order 5.

Regarding the importance of curriculum domains in order of rank, the results shown in Table 3 demonstrate that the Illinois and Florida teacher groups tended to rate the curriculum domains more similarly than the Professors of Curriculum. These findings might be a reflection that teachers work with colleagues who are involved in similar kinds of activities, roles and behaviors on an everyday basis. Irrespective of the ranking of domains by the Illinois and Florida teachers and Professors of Curriculum it is notable that every domain evidenced a high level of homogeneity, with alphas ranging from .6236 to .9332.

The nine domains represent the broad areas of knowledge important to the field of curriculum and suggest what curriculum specialists should know about the field. The 47 curriculum practices categorized within the domains represent important activities that describe what curriculum specialists do. Together the curriculum domains and curriculum practices represent the knowledge base of the field and a partial compendium of behaviors that curriculum specialists engage in while inquiring about planning an d implementing the curriculum.

Conclusions

By themselves, the 55 curriculum practices represent the important behaviors of curriculum specialists. Although no educational program can be devised which will encompass all agreed upon knowledge, it is essential to determine and operationalize what practices are needed to improve the curriculum process. In order to engage in dialogue or inquiry about curriculum domains, it is important that these constructs be defined in the same way. It seems that empirical investigations are needed to clarify domains if we hope to move discussions beyond the linguistic and metaphorical levels.

The behaviors and activities listed in Table 1 help establish behaviors or criteria for the emerging roles of the curriculum specialist. This compendium of practices might have utility as an evaluation tool for principals. Principals might use this list to assess teachers' instructional skills and identify their methodological strengths and weaknesses. The curriculum practices serve as criteria or requirements for graduate study involving curriculum certification, staff development for curriculum specialists, and for making curriculum decisions from many levels—school, district, and community.

I have postulated that the curriculum practices identified are representative of the kinds of behaviors that curriculum specialists engage in or perform. Most important, they are measurable and observable behaviors for theorists and practitioners to study, and possibly use for assessment in school settings.

As the field of curriculum seeks to identify a compendium of operational roles, this quantifiable knowledge base of 9 domains, and 47 practices might be helpful in defining what curriculum specialists should know and be able to do. Analyzing the frequency and conditions under which thee behavior scan be observed in real situations degree to which they are emphasized in schools and classrooms might extend our understanding of the empirical relationships between theory and practice and promote successful implementation of school improvement processes.

I have made certain assumptions perhaps controversial in nature, **domains** represent the broad content areas that practitioners should know and be able to utilize in actual situation and **practices** refer to the specific roles of curriculum specialists and supervisors. In order to implement successful school improvement action plans, there will need to be agreement concerning domains and practices so that objective and quantifiable criteria can be established. Currently many curriculum processes and decisions are made in nonconsensual ways. An agreed upon set of domains and practices should benefit the field of curriculum.

This study is an attempt to establish an empirical format for identifying **curriculum domains** (the knowledge base or important content areas of the field) and **curriculum practices** (precise activities curriculum specialists perform). Identifying the important content areas in curriculum, is essential to specifying the kinds of skills and behaviors that curriculum workers should acquire as a result of their postsecondary education. Postsecondary education departments and colleges of education should take a leadership role in establishing criteria that define professional standards. These programs should implement accountability measures that ensure that graduates have acquired the necessary skills for real world applications of curriculum. By creating standards for professional practice and producing competent curriculum workers, postsecondary programs of education can play an integral role in helping schools to successfully implement school improvement initiatives.

REFERENCES

American Heritage Dictionary (1983). New York, New York: Dell Publishing Company.

Beauchamp. G. (1961). *Curriculum Theory*. Wilmette, IL: Kagg Press.

Behar, L. (1992). *A Study of Domains and Subsystems in the Most Influential Textbooks in the Field of Curriculum 1970-1990).* Unpublished doctoral dissertation, Loyola University of Chicago.

Doll, R.C. (1989). *Curriculum Improvement: Decision Making and Process. Seventh Edition.* Boston, Massachusetts: Allyn and Bacon.

English, F.W., Ed. (1983). *Fundamental Curriculum Decisions.* Alexandria, Virginia: Association for Supervision and Curriculum Development.

Gideonse, H.D. (1989). *Relating Knowledge to Teacher Education: Responding to NCATE'S Knowledge Base and Related Standards.* New York, New York: American Association of Colleges for Teacher Education.

Giroux, H.A., Penna, A.N. & Pinar, W.F. (1981). *Curriculum & Instruction: Alternatives in Education.* Berkeley, California: McCutchan.

Gudmundsdottir, S. (1991). Values in Pedagogical Content Knowledge. *Journal of Teacher Education. 41*(3), 44-52.

Kimpston, R.D. and Rogers, K.B. (1986). A Framework for Curriculum Research. *Curriculum Inquiry. 16*(4), 463-474.

Madaus, F. and Stufflebeam, D. Editors. (1988). *Educational Evaluation: Classic Works of Ralph W. Tyler.* Norwell, Massachusetts: Kluwer Academics.

MacDonald, J.B., Ed. (1965). *Theories of Instruction.* Washington, D.C.: Association for Supervision and Curriculum Development.

McNeil, J.D. (1990). *Curriculum: A Comprehensive Introduction.* Fourth Edition. Glenview, Illinois: Scott, Foresman.

Ornstein, A.C. and Hunkins, F.P. (1993). *Curriculum: Foundations, Principles and Issues.* Second Edition. Boston, Massachusetts: Allyn and Bacon.

Ornstein, A.C. and Levine, D. (1989). *Foundations of Education.* Fourth Edition. Boston, Massachusetts: Houghton Mifflin.

Ornstein, A.C. (1987). The Field of Curriculum. What Approach? What Definition? *High School Journal. 70*(4), 208-216.

Rosales-Dordelly, C.L. and Short, E. (1985). *Curriculum Professors Specialized Knowledge.* Lanham, Maryland: University of America Press.

Saylor, J.G., Alexander, W.M. & Lewis, A.J. (1981). *Curriculum Planning for Better Teaching and Learning.* Fourth Edition, New York, New York: Holt, Rinehart, & Winston.

Schubert, W.H. (1986). *Curriculum: Perspectives and Practice.* New York, New York: MacMillan.

Scriven, M. (1967). Goal-free evaluation. In E. House (Ed.), *School Evaluation.* Berkeley, California: McCutchan.

Stufflebeam, D., Foley, W. Gephart, W. Guba, E., Hammond, R., Merriman, H. and Provus, M. (1971). *Educational Evaluation and Decision Making.* Itasca, Illinois: F.E. Peacock.

Tyler, R.W. (1949). *Basic Principles of Curriculum and Instruction.* Chicago, Illinois: University of Chicago Press.

Zais, R.S. (1976). *Curriculum: Principles and Foundations.* New York, New York: Thomas Y. Crowell Company, Inc.

Assessing Approaches to Classroom Assessment: Building a Knowledge/Skill Base for Preservice and Inservice Teachers

Lehman W. Barnes, Ph.D.
Marianne B. Barnes, Ph.D.
Division of Curriculum and Instruction
College of Education and Human Services
University of North Florida
4567 St. Johns Bluff Road
Jacksonville, FL 32224

Assessing Approaches to
Classroom Assessment:
Building a Knowledge/Skill Base for Preservice and Inservice Teachers

Rationale:

In the past several decades a number of elementary science curriculum projects have been conceptualized, tested, and presented to school leaders and teachers (Cain and Evans, 1984). These projects have met with varying degrees of success due, in part, to a variety of variables operating in the school setting. One crucial variable has been the involvement of the teacher in the development process—from conceptualization and initial involvement to development and trialing, to debugging and critiquing, to adaptation and to integration into his/her particular school setting. In interviews and conversations with many K-12 teachers, we have become acutely aware of what the research literature suggests - that ownership and meaningfulness are crucial in the development of an idea, a strategy, and approach, or a curriculum (Betkouski, 1989). Many sources of expertise should be tapped as we plan and evaluate teaching, learning and assessing strategies; the teacher must be an integral and continuing part of this process (Worthen, 1993). The process itself should begin in methodology courses by supplying opportunities which require students to experience and critique various assessment modes. Exposure to and reflection upon assessment approaches should precede teacher implementation of such approaches for purposes of school improvement.

Research in learning is moving in the direction of analyzing how the mind makes connections which are activated in the presence of new information. Learning science involves personalization of individual "stories" (Martin and Brouwer, 1991). In addition, context affects the learning process (Gardner, 1991), as does the individual's ability to fully immerse the self in engaging activities (Csikszentmihalyi, 1990). This line of research presages a holistic approach to teaching and to teacher training which is far more experiential than current programs encourage. One might hypothesize that methods students who are asked to design classroom experiences and assessments which are varied and experiential should themselves engage in like activities in conducive settings. The studies described are attempts to begin to determine effective ways to teach meaningful assessment approaches to pre- and inservice teachers.

Study 1 Sample:

Thirteen students enrolled in a secondary science methods class, eighteen students enrolled in a combined elementary/secondary science methods section, and twenty-three minority teachers enrolled in a Science Technology, and Society course within an elementary master's degree program were involved in the study. All were students in programs at a southeastern state university during the 1992-92 academic year. Instructor 1 taught the elementary masters group, while Instructor 2 taught the other two groups. The "alternative assessment" group (combined elementary/secondary methods section) taught by Instructor 2 experienced a course based primarily on a variety of alternative assessments, including keeping a journal.

Procedures:

Data were gathered by Instructors 1 and 2 on attitudes toward various modes of science assessment. The secondary group was "walked through" a density process assessment as a group activity, concluding that they just look to their own goals and objectives in the design and scoring of assessment. Then they were tested on the first twelve items of the General Science Test - Level II (Australian Council for Educational Research), providing answers, reasons for the answers, and a brief description of an alternative assessment, provided that they believed that the knowledge required in the item could be asked to choose one item from the General Science test which "inspired" them to inquire further into the topic. They were to describe the design of the inquiry and an associated assessment procedure.

The combined elementary/secondary methods class formed groups of 2-3 people and were assigned one item of the General Science test. They were to answer it and provide a reason for the answer. They addressed the following:

- Which instructional objective(s) do you think this item attempts to assess?

- Is the test item a "good" item? Why or why not?
- Design an alternative, inquiry-oriented method to assess this or a related objective(s)?
- Compare the test item provided to the alternative method which you devised in terms of information provided to the student, assessment information provided to the instructor, and perceived level of interest by a student.
- List several goals of classroom assessments of science knowledge and skills.
- Based on these goals, list and describe several kinds of assessment methods which you would use in teaching science.

The STS elementary teachers were administered the Level 1 General Science Test (Australian Council for Educational Research). For selected test items, the teachers engaged individually in a process of responding to the item, the teachers engaged individually in a process of responding to the item, giving reasons for their responses, doing an item critique, and suggesting an alternative assessment. The teachers constructed alternative assessments and compared these alternatives to the original paper-pencil assessment item. After completion of this general sequence, teachers were asked to design alternative assessment for a science topic/concept/activity of their choice. Sample activities were present by individual teachers to their colleagues. Group discussion ensued concerning the impact of the assessment experience both on their knowledge of and previous experience with alternative assessment strategies.

Results and Discussion of Study 1:

Preferred assessments on the part of all subjects are summarized as follows:

PREFERRED ASSESSMENTS IN STS COURSE TAUGHT BY INSTRUCTOR 1 (N = 23)

Practical, Hands-On	5
Multiple-Choice	18
Essay	7
Demonstration	2
Oral Discourse	2
Projects	6
Experiments	2
Take-Home	1
True-False	3
Fill in the Blank	4

Preferred Assessments in Science Methods Courses Taught by Instructor 2

Alternative Assessment Group		Traditional Assessment Group
	N = 18	N = 13
Observation	1	
Journals	3	
Practical, Hands-On	8	7
Application	2	
Multiple-Choice	3	3
Essay	3	3
Demonstration	1	
Oral Discourse	1	
Projects	1	
Problems	1	2
Matching	1	

Paper-Pencil		3
Variety	1	1
No answer		2

A question which emerged from the data analysis is as follows: DOES EXPOSURE TO ALTERNATIVE ASSESSMENT ALTER ASSESSMENT PREFERENCE?

From the activity in which nine groups of students in the combined elementary/secondary course designed alternative assessments for assigned test items, the following characteristics which described instructional foci were noted:

Use materials	4
Employ small groups	2
Classify	1
Predict	1
Make Observational Record	1
Hypothesize	1
Use/Build Model	3
Present - Written/Oral	1
Compute	1
Do Background Research	2

A comparison of these characteristics to the original assessment preferences expressed revealed greater specificity and emphasis on the process skills - the doing of science. In addition, less emphasis is placed on the more traditional modes of assessment. At the conclusion of the experimental exercise, the following list of assessment methods which they would use in teaching science was induced from the group responses:

Journals/logs	3
Projects	3
Tests	5
Cooperative projects	3
Practical/labs	3
Observation and recording	3
Reports	
Oral	2
Written	2
Free responses tests	1
Scientific article critiques	1
Field trip with report	1
Personal conferences	1
Problem-solving situations	1

Examination of these responses poses a question for further investigation: IS ASSESSMENT VARIETY ENHANCED BY A SPECIFIC TREATMENT DESIGNED TO HEIGHTEN AWARENESS?

In the secondary methods group taught by Instructor 2, the following approaches to design of alternative activities suggested by topics covered in standardized test items emerged:

(N=13 secondary science methods students)
Form hypotheses
Collect data and decide
Examine maps
Use group observations, activities, reports
Build models (inc. computer)
Use props, materials
Vary the context

Visit field sites
Construct apparatus
Allow student decision-making in design; explore
Construct graphs

Initially the STS elementary teachers expressed a variety of preferences for personal science assessment and, for the most part, these preferences tended toward the typical paper-pencil format, e.g., multiple-choice, essay, true-false, fill in the blank. Some did mention projects but these were considered primarily as library research type projects. During the post assessment sequence group discussions, most participants mentioned an interest in and positive response to assessment strategies which involved the student in some kind of hands-on performance. Their designed assessment activities reflected this positive response in that most (but not all) of the sample student activities required some degree of student hands-on performance.

Sample Student Assessment Activities Designed by Teachers

- density activity - student compares fresh water, salt water, oil, glycerin
- field activity - student constructs map and uses compass in field study
- physical/chemical change activity - student demonstrates examples
- weighing activity - student weighs dry and wet items and computes weight change and compares relative absorptions
- earth worksheet activity - student reads and completes work sheet
- sink/float activity - student classifies objects as "sinkers/floaters"
- ocean/water cycle activity - student responds to yes/no questions
- inclined plane activity - student demonstrates variation in the distance an object rolls relative to inclination of plane
- measurement activity - student measures various size rocks using metric ruler and tape
- weather instruments activity - student constructs anemometer and weather vane and demonstrates data collection and analysis
- water bottle activity - student sets up bottles with varying amounts of water and demonstrates variation in pitch
- magnet activity - student classifies objects as attracted to or not attracted to magnets
- probability activity - student demonstrates and explains likelihood of color selection when a 2 red/1 blue ratio exists
- spring scale activity - student demonstrates relationship of mass of object to force required to lift object
- solution activity - student makes own solution and identifies solute and solvent
- volume activity - student demonstrates and computes relative volumes of different sized containers
- boiling water activity - student demonstrates variation in time to initial boiling related to volume of water in container
- multisurface activity - student orders objects in accordance with numbers of surfaces
- making inferences activity - student responds to questions after reading a paragraph concerning energy
- rock measurement activity - student makes measurements of rocks in reference to length and mass

Study 2 Sample:

Fifty-six students enrolled in two sections of undergraduate elementary science methods and thirteen students enrolled in graduate elementary science methods during the spring 1993 term at a southeastern state university will be involved in a further study building upon the methodology outlined below.

The main focus of Study 2 was the degree of transferability of learning how to assess one concept in a variety of ways to assessing other concepts. Thus the question posed is:

DOES THE EXPOSURE TO ASSESSMENT VARIETY TRANSFER FROM AN INITIAL TREATMENT OF A PARTICULAR CONCEPT TO THE EMPLOYMENT OF AN ASSESSMENT VARIETY WHEN DEALING WITH OTHER CONCEPTS?

Procedures for Study

Design within the two sections of elementary science methods:

Both sections received similar instruction concerning the topic of density. Instruction included a video scenario, a density inventory, practical hands-on activities and lecture/demonstration/discussion. The two sections differed in two respects:

Section A - Instructor emphasized the interrelatedness of instructional activities and assessment activities, i.e., that any activity could function as an instructional activity and/or an assessment activity.

The students were given a multiphasic assessment at the end of their density experiences. This assessment included paper/pencil and practical components.

Section B - Instructor did not emphasize the interrelatedness of instructional activities and assessment activities.

The students were given a paper/pencil assessment at the end of their density experiences.

Following the density assessment, both sections received similar instruction concerning the topic os simple machines with emphasis on levers and inclined planes. Students participated in practical activities, paper/pencil problem-solving activities, and lecture/demonstration/discussion. At the conclusion of this instruction, each section was asked to design instructional activities and an assessment for a mini-unit on levers and inclined planes.

Results and Discussion of Study 2 - elementary methods:

Section A demonstrated a greater variety of both instructional and assessment components and strategies than did Section B. In terms of assessment variety, the following approaches were most abundant:

Section A	Section B
paper/pencil questions	paper/pencil questions
paper/pencil problems	paper/pencil problems
practical questions	
practical problems	
video scenario problems	

The greater variety of assessment strategies created by Section A might be the result of a simple student modeling of instructor behavior. Note that the students in both sections did receive a variety of instructional strategies but differed in exposure to assessment strategies and emphasis on the interrelatedness of instruction and assessment. The results suggest that the exposure to assessment variety with one concept may transfer to employment of assessment variety with other concepts.

Design within Graduate section of Elementary Science Methods:

Students were asked to brainstorm and reflect upon a number of reasons for assessment. They addressed aspects which affected students, teachers, parents, and administrators and reinforced the linkage between instruction and assessment. The list was embellished with further reading from the course text. An inventory assessing the concept of density was then administered; it required a true-false answer and the choice of a reason for the answer. The students had already received instruction about density by means of a short video scenario, The Golden Statuette, as a preassessment and hands-on activities using metal bars, a double pan balance, and various liquids. After discussion of the inventory and its advantages and disadvantages, the students planned further approaches to assessing student knowledge of density: e.g., group and individual projects, portfolios, writing about the concept, videos and videodiscs, and using hands-on activities.

Following a break, the problem "What causes a hot air balloon to rise?" was asked and followed by a videodisc demonstration interspersed with practical activities and a story; density activities (of solids, liquids, and gases) were demonstrated and discussed with respect to both instruction and assessment. Student reaction to the entire demonstration was highly positive.

The students in the methods course were assessed by means other than paper and pencil tests: written assignments, an individualized project, an earth month project which could be planned in pairs; and the preparation of a 5-day sequence of lessons with assessments. The students presented their lessons and assessments to the class.

From their prepared packages of lessons and assessments, the kinds of assessments were determined. Findings are as follows (* = number of students):

- preassessment through classroom discussion*
- design or create something individually (plant mobile, story, volcano)****
- assemble something*
- write words in book*
- draw and label****
- trace*
- oral report**
- name or locate something***
- give examples*
- written report******
- group creation (story, poem, play, soup, garden layout)****
- cooperative group worksheets*
- describe and summarize (vague)*
- traditional written test******
- pre and post free writing*
- computer game*
- journal****
- make chart or data sheet***
- make list**
- do a calculation*
- teacher observation (vague)******
- checklist**
- responses to oral questions**
- student self-evaluation***
- videotape students*
- hands-on assessment centers with worksheets*
- student portfolio**
- role play*

Previous studies by the researchers on a process assessment of the concept of density have underscored the importance of allowing for individual differences in demonstrating and verbalizing responses to the test items. Preliminary work on involving preservice teachers in critiquing assessment strategies had been conducted by the researchers i methodology classes during the 1989-90 and 1990-91 academic years. The current studies build on these findings by creating settings in which teachers can engage fully and personally in the assessment process. By taking, critiquing and constructing assessment items, teachers were exposed to the complexity of the assessment process and their need to assume responsibility for choosing appropriate assessment protocols which will reflect authentically their own students' knowledge and skills and expand their own concept of assessment. By choosing assessment approaches within varying contexts, students built upon their own course experiences with the concept of density with fair to very good degrees of success. Instructional activity and assessment should be equivalent constructs with both supporting a tone of inquiry within a methods course. An instructional sequence for use in science methods classes might be that which is depicted in the attachment. In addition, thoroughly assessing a single concept as a class set of activities appears to positively affect the variety of assessments chosen by the students in designing their own lesson sequences.

REFERENCES

Betkouski, M. (1989). *Meaningful learning in elementary science:* Dubuque: Kendall/Hunt.

Cain, S.E., Evans, J.M. (1984). *Sciencing - an involvement approach to elementary science methods.* Columbus: Charles E. Merrill Publishing Co.

Csikszentmihalyi, M. (1990). *Flow - the psychology of optimal experience.* New York: Harper and Row.

Gardner, H. (1991). *The unschooled mind - how children think and how schools should teach.* Basic Books.

Martin, B.E., Brouwer, W. (1991). The sharing of personal science and the narrative element in science education. *Science Education.* 75 (6), 707-722.

Worthen, B.R. (1993). Critical issues that determine the future of alternative assessment. *Phi Delta Kappan.* 74 (6), 444-454.

A SEQUENCE:

PAPER-PENCIL ASSESSMENT ITEM → PERSONAL RESPONSE → REASONS FOR RESPONSE → ITEM CRITIQUE → SUGGEST ALTERNATIVE ASSESSMENT

→ CONSTRUCT AND TRIAL ASSESSMENT → COMPARE TO PAPER-PENCIL ASSESSMENT → REVISE ALTERNATIVE ASSESSMENT

Assessing Mathematical Problem Solving from Multiple Perspectives

Mary Grace Kantowski
Stephanie Robinson
Thomasenia Adams
Juli Dixon
Jeff Issacson

Presented at the William F. Breivogel Annual Conference
Florida Education Research Council
March 25, 1993

Mathematical education has been in a gradual yet constant state of change. *The Curriculum and Evaluation Standards for School Mathematics* by the National Council of Teachers of Mathematics (NCTM, 1989) provide the goals and objectives needed to guide these changes. In reality, changes in instruction and curriculum will be driven by changes in assessment and evaluation.

The object of school mathematics is the development of mathematical power for all students (NCTM, 1989). Mathematics provides the opportunity for children to learn the power of thought as opposed to the power of authority, and to develop critical habits of mind. "Inherent in this document is a consensus that all students need to learn more, and often different, mathematics and that instruction in mathematics must be significantly revised" (NCTM, 1989). However, change is a process, not an event. The characteristics of change are subtle and complex. For change in instructional goals to be effective, change must also occur in textbooks and assessment process. Teachers teach what is in the textbook and students learn what will be on the test. Therefore, texts should emphasize and tests must measure what is most important. ". . . we must ensure that tests measure what is of value, not just what is easy to test. If we want students to investigate, explore, and discover, assessment must not measure just mimicry mathematics" (National Research Council, 1989).

Changes in assessment must include alternative techniques as well as improved paper and pencil tests: a complete assessment system, formal and informal. Assessments of students' mathematical learning must coincide with daily classroom practices and be directly related to instruction (Sammons, Kobett, Heis, & Fennell, 1992) and goals for student learning. Assessment techniques must be used to evaluate process and product, if that is to be the focus of instruction.

Assessment should be ongoing, as learning is ongoing. Assessment should be active and constructive, as learning is active and constructive. Assessment should involve a variety of methods and tasks, as learning involves a variety of methods and tasks. Evaluation information should be used to make decisions about the content and methods of instruction and about the classroom environment, to communicate what is important, and, as traditionally used, to assign grades (Jawojewski, 1991). Consistent with the *Standards*, instruction and evaluation should focus on students' abilities in problem solving, reasoning, and communicating mathematically, and students' disposition toward mathematics. Evaluation standards promoted by NCTM correspond to the standard for instruction and curriculum. *The Professional Standards for Teaching* (NCTM, 1991) also reflect the emphasis on, and coordination of, instruction and evaluation.

Traditional forms of assessment have focused primarily on paper and pencil tests: multiple-choice, short answer, fill-in-the-blank with a single right answer with every student working alone. This produces information about the quantity of student knowledge, but not necessarily the quality of student knowledge, nor the processes being used.

Evaluation techniques must change to reflect changes in instructional methods, modes, and goals. Multiple means of assessment, including written, oral and demonstrative formats, are needed to provide a complete view of student learning. Teachers gather informal assessment information continuously during instruction; however this information lacks structure and documentation. A variety of techniques and modes can be used to provide this information (Clarke, et al, 1990; NCTM 1989; Stenmark, 1991). These methods include:

open-ended questions, discussion questions
structured or open-ended interviews
homework
individual or group projects
journals and essays
class presentations and dramatizations
computer and calculator activities
small group activities
observations and anecdotal records
portfolios
performance and exploratory activities
holistic scoring
student self-evaluations, peer evaluations

Some of these ideas are self-explanatory, others need further discussion. For all, however, teachers need to plan, prepare, and practice to be effective.

Observations may be planned or unplanned. The teacher may list key questions or targeted behaviors to be assessed. Specific statements or actions kept as anecdotal records supplement observations and provide specific feedback and information. The statements may be kept on index cards, checklists, or in portfolios. Observations can provide insights on diagnoses of processes and on students' abilities to solve problems, communicate, and cooperate. Many components of the student's disposition toward mathematics can be seen in behavior which indicates confidence, flexibility, perseverance, curiosity, reflection, an appreciation of the value of mathematics. Interviews and conferences, which may be structured or open-ended, can complement this information.

Students can demonstrate knowledge of a concept or skill by completing a mathematical task or activity, individually or in small groups, as classwork or as part of a test. Some activities may include modeling a mathematical concept, preparing class presentations and demonstrations, conducting experiments, completing multicurricular (integrated) assignments, and participating in exploratory and problem-solving activities. Data analysis projects are a rich source for problem solving: collecting data, organizing, classifying, interpreting, making conjectures and supporting those conjectures are excellent examples of meaningful mathematical activities.

Small group activities allow the teacher to observe the extent to which the student is involved, cooperates, and communicates with peers. Grades can be assigned to the whole group, the individual, or a combination. For example, the group can participate in the activity, then each student can write a description, solution, and analysis of the task and its results (Coxford, 1991).

Self-assessment or peer assessment is also a useful evaluation tool. Use of a checklist, or daily journal entry, may facilitate this method. Joint student-teacher evaluations can be the basis for an interview or conference.

The purpose of the following set of problems is to explore a geometric transformation (called rotation) through conjecturing and the *Geometer's Sketchpad* (geometric software for the Macintosh), while illustrating multiple forms of assessment including many of those previously discussed. Informally, a conjecture is an "educated guess" or hypothesis believed to be true about all objects of the type in question. A conjecture may, in fact, turn out to be false. For this reason, external assessment will, in part, be based on the reasonableness of the conjecture will be externally assessed for valid mathematical reasoning. Self (internal) assessment will take lace as the student tests the conjectures and describes the consequences in writing. In this way, the teacher has a sample of the student's reasoning dealing with rotations, the student has the chance to articulate his/her ideas through multiple representations (pictures and writing), and the student enjoys immediate feedback based in his/her conjecture.

> Directions: For each problem, use the highlighter pen to record your conjecture on the diagram. Use an ordinary pen or pencil to record the actual result of the transformation on the same diagram. In the journal area, include the reasoning behind your conjecture and the reaction to the consequences of testing the conjecture on the *Geometer's Sketchpad*. In each of the problems, the center of rotation has been marked on the diagram (see attached activity sheet).
>
> Problems:
>
> Rotate the flag 60° about the marked center.
> Rotate the flag -45° about the marked center.
> Rotate the flag 135° about the marked center.

Paper and pencil tests judge some aspects of students' knowledge; however, these, too, can be improved to include factors not readily available with traditional tests. Open-ended questions allow the teacher to view the student's process as well as final product. Questions may have more than one answer, more than one solution process, and more than one interpretation. Many typical textbook and test questions can be modified to be open-ended (Schulman, 1992). Revisions can include:

- written stories to fit data, graphs, lists, new questions regarding the same information or extensions of the original question,
- discussions of the approach taken to understand the problem or to solve the problem,
- discussion of 'why' and 'how' and 'explain,'
- reversal of the given data and required information,
- missing or unknown information, and,
- nongoal specific tasks.

Tests may also be student-created or involve completing a practical or exploratory activity.

In many cases, one problem may be used as a paper and pencil question or be rich enough to be the basis of a small group of whole class discussion. One question may have a variety of answers depending on the conditions and variables, as well as the perspective of the student. The following problem may be a single answer question on probability, or a full discussion on the results as conditions and variables change (The full activity sheets are attached).

You Decide

Ms. Smith hired Mr. Jones to work at the Lose-It-All Casino. Ms. Smith made the following salary offer to Mr. Jones. He can receive $400 per week, or he can determine his weekly salary by the following game of chance:

Draw three bills out of a bag containing six bills. The bag contains a $5, a $10, a $20, a $50, a $100, and a $500 bill. The value of the three bills will be Mr. Jones weekly salary for the year.
Which offer should Mr. Jones accept?

Before answering this question, let's look at a simpler question. Suppose Mr. Jones were to draw two bills out of a bag containing a $10, a $50, a $100, and a $500 bill. We could list all the possible outcomes:

$500 - $100
$100 - $ 50
$500 - $50
$100 - $20
$500 - $20
$ 50 - $20

Half the outcomes are greater than $400, and half the outcomes are below $400. Does this help Mr. Jones make a decision?

Let's analyze this further. The *expected value* is the average amount of money you would draw from the bag each time if you played the game many, many times. We find the expected value by multiplying the value of each outcome times the probability of that outcome; then adding all these products together. Try it for the four-bill game. The probability for each draw is ${}^{1}/_{6}(\$500 + \$100) + {}^{1}/_{6}(\$500 + \$50) + {}^{1}/_{6}(\$500 + \$20) +$
${}^{1}/_{6}(\$100 + \$50) + {}^{1}/_{6}(\$100 + \$20) + {}^{1}/_{6}(\$50 + \$20) = \$335$

We can now return to the six-bill game.

First, list all the possible outcomes when drawing three bills from the bag (see the full activity).
How many different outcomes are possible?
What is the probability of each outcome?
What is the probability of drawing more than $400?
What is the expected value?

Based on the probability and expected value, should Mr. Jones take the $400 salary or should he play the game and let the draw determine his salary? Are there any other factors that Mr. Jones should consider before making this decision? What if Mr. Jones had the option to draw every week?

Student perceptions and interpretations of the problem are important factors in paper and pencil questions. The perceptions and interpretations may be due to cultural, societal, environmental, or mathematical background differences as well as a combination of the aforementioned. What is considered "real world" may in fact be subjective. Compare the following problems taken from the Second International Mathematics Study:

1. A car take 15 minutes to travel 10 kilometers. What is the speed of the car?
2. A freight train traveling at 50 kilometers per hour leaves a station 3 hours before an express train which travels in the same direction at 90 kilometers per hour. How many hours will it take the express train to overtake the freight train?
3. The graph shows the distance traveled by a tractor during a period of 4 hours. How fast is the tractor moving? (The graph is a linear progression in terms of distance and time.)

To answer all problems "correctly" requires that the student assumes that the car or train or tractor travels at a constant rate of speed. The first question makes that easier to assume than the other two. A car already in progress may maintain the same rate of speed for fifteen minutes. However, a train leaving a station does not travel at a constant rate. It must start off more slowly and pick up speed. Also trains may vary their speed based on the conditions of the track, terrain, and whether crossing roads or passing through cities. The question concerning the tractor may also involve interpretation. Is the tractor on a highway or in the field? In either situation, the rate of speed may not remain the same. Traffic conditions on the road may need to be considered. Does the tractor have to pull off periodically to let cars pass? In the field, is the ground consistency the same? Does the tractor make passes up and down the field, turning around often. Other factors are involved and the "correct" answer may be "not enough information." The question, therefore, is not necessarily realistic to all students. Traditional paper and pencil problems may not have a single right answer. Multiple-choice tests cannot reflect the thought processes of the student. A student's thought process may lead to conflict with the "correct" answer, but may, in fact, be more legitimate.

Student portfolios can contain written tests, problem solutions, a well as notes on investigations, oral reports and responses to open-ended problems. The teacher's observations, checklists, and interview notes may also be included. Another option is for the portfolio to be created by the student, as an artist's portfolio, with samples of specific types of things.

Journals entries about classwork, homework, activities, and projects are a rich source of information. If students are encouraged or required to record entries in journals, then the writings should be included in the assessment process. This reinforce the importance of communicating in mathematics. Steen (1992) proposed that all students should learn to communicate mathematically. Communication in this sense does not differ form our well-known means of communication: speaking, listening, reading, and writing. NCTM (1989) combines these methods of communication in mathematical language with reflection and demonstration . In general, "... basic communication skills are more important than ever before since they are both contributing to and necessitating a literacy-intensive society" (Romberg, 1990, p. 469). In the mathematics classroom, the implementation of journal writing can enhance students' experiences with communicating in mathematics.

Journal writing provides the means by which students can foster the need for reflecting on and assessing their own understanding (Lester & Kroll, 1991). Clarke (1992) posited that students' reflection on their own learning is an inherent benefit. Moreover, teachers can use students' journal entries to monitor students' mathematical understanding (Collison, 1992). In addition, writings can be used to determine whether or not students have assimilated and accommodated the mathematics and if they are able to manipulate the mathematics for purposes of application and communication (NCTM, 1989).

There are no specific guidelines for implementing journal writing. The teacher can provide topics from which students can choose to write about or students can have the freedom to write about personally chosen topics. The subject can be directly related to the mathematical topic or can be initiated to collect information on the students beliefs and attitude in mathematics (Collison, 1992). The following list provides some indication of the types of topics that students can be instructed to write about:

1. What is your favorite mathematical topic? Explain.
2. What is different about the mathematics you learned last year and the mathematics you are learning this year?
3. Suppose you could speak with a famous mathematician. Who would you speak with? Why? What would the conversation be like? What kinds of questions would you ask her/him.
4. How would you convince a doubting classmate that mathematics is a useful discipline?
5. Can you connect the mathematical learning that you encountered today with any learning of the past?

Whether the student or the teacher chooses the topic of the journal entry is of little consequence. Student-chosen topics provide the teacher with some indication of the ideas that students value or choose to emphasize. In addition, if the student is not restricted by teacher-chosen topics, then the student might be more apt to discuss their feelings about mathematics and their learning of mathematics which he/she might not otherwise discuss. Teacher-chosen topics provide guidelines which keep students on a common task; thus, the writings can be used to examine the level of all students on one topic. However, this does not indicate that the writings should be used to compare students for the purpose of assigning grades, but should be used for the purpose of arranging instruction to meet the various needs of the students.

Since a consequence of journal writing is that the journals must be legible and comprehensible, students are encouraged to improve not only communication in mathematics, but communication skills in the English language. The teacher of mathematics must be prepared to assist students with applying appropriate writing techniques. With all the other responsibilities that are taken on by the classroom teacher, this may seem an added burden. However, if assessment techniques in mathematics are truly going to be alternative, then teachers of mathematics must be prepared to evaluate the information gathered in such a way that learning and instruction can be improved.

The final activity reflects some of the insights mentioned above on journal entries. The activity is to actually write a journal entry for this day's session. When implementing assessment through journal writing, one of the issues that must be addressed by the teacher is the purpose of the journal entry. Teachers must convey to students what a writing assignment is intended to reveal. For instance, if the teacher really is interested in students' reflective thinking after their participation in a mathematical activity or after completing a project or assignment (which is the focus of the entry today), then the following could be proposed to the students:

> For today's journal entry, please respond to the following questions:
>
> 1. Before participating in the current activity, what kinds of strengths or weaknesses affected the role you chose to play in the activity?
> 2. If you could have a different role, what would it have been? Why?
> 3. During your participation in the activity, how do you think what you did or didn't learn affected your academic potential?
> 4. What made the activity interesting and/or worthwhile to you?
> 5. How would you change the activity to better benefit you and/or other students?
> 6. How did your participation in the activity make you feel about your learning of mathematics?

However, if the teacher is more interested in students' evaluation of an activity than the students' reflection on an activity, then the following questions could be posed:

> 1. Explain how the current activity has helped you to grow mathematically.
> 2. What suggestion(s) can you make which might improve the presentation of the activity?
> 3. What type of resources could you suggest that might be used to design a better presentation?
> 4. Sometimes, we feel that "something was missing" from an activity. Did you feel that way while participating in the activity? If so, can you suggest what "was missing?"

Teachers who implement journal writing activities may perhaps be able to construct other questions or activities which might prompt students to write about different topics. In any case, it should be noted that when students write in their journals, they often take the opportunity to deviate from the instructions given by the teacher. Thus the instructions must be clear and straightforward as to encourage students to provide a response that is appropriate. There are many purposes and many subjects for journal entries. They can be an important part of the overall student portfolio for evaluation.

A major goal of assessment is the improvement of instruction and learning. As one improves, so must the other. The vision for the future of mathematics education is for all students to gain mathematical power by doing mathematics, being active on real problems in a curriculum that continuously evaluates teacher, student and content for improvement. The joint effort of the entire educational community including every teacher actively involved is required to effect change. The focus of the present and the future is to actively make changes in instruction and assessment to reflect the changes in knowledge about the learning process, changes in the curriculum, and changes in the world in which we live.

REFERENCES

Clarke, D.J. (1992). Activating assessment alternatives in mathematics. Arithmetic Teacher (Focus Issue on Assessment): 24-29, Feb.

Clarke, D.J., Clarke, D.M., & Lovitt, C.J. (1990). "Changes in mathematics teaching call for assessment alternatives." In Teaching and learning mathematics in the 1990's. Reston, VA: NCTM.

Collison, J. (1992). Using performance assessment to determine mathematical dispositions. Arithmetic Teacher, (Focus Issue on Assessment): 40-47, Feb.

Coxford, A.F. 91991). Geometry from multiple perspectives. Reston, VA: NCTM.

Jawojewski, Judith S. (1991). Dealing with data and chance. Reston, VA: NCTM.

Lester, Jr., F.K. & Kroll, D.L. (1991). Evaluation: A new vision. Mathematics Teacher (April): 276-83.

National Council of Teachers of Mathematics (1989). Curriculum and Evaluation Standards for School Mathematics Reston, VA: NCTM.

National Council of Teachers of Mathematics (1991). Professional Standards for Teaching Mathematics. Reston, VA: NCTM.

National Research Council (1989). Everybody Counts.

Romberg, T.A. (1990). Evidence which supports NCTM's Curriculum and Evaluation Standards for School Mathematics. School Science and Mathematics, 90(6): 466-79, Oct.

Sammons, K.B., Kobett, B., Heis, J; and Fennell, F. (1992). Linking instruction and assessment in the mathematics classroom. Arithmetic Teacher (Focus Issue on Assessment): 11-16, Feb.

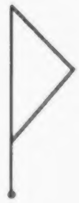
Schulman, Linda (1992). Centerpiece. Mathematics-Science-Technology News from Lesley College. Winter, 1992.

Steen, L.A. (1992). Does everybody need to study algebra. Basic Education, 9-12, Dec.

Stenmark, J.K. (Ed.) (1991). Mathematics assessment: Myths, models, good questions, and practical suggestions. Reston, VA: NCTM.

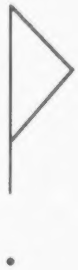
Description of transformation:
Rotate the flag 60° about the
marked center.

If your conjecture is different from the computer-generated rotation, explain why.



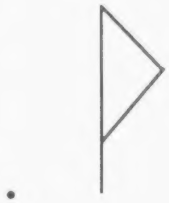
Description of transformation:
Rotate the flag -45° about the
marked center.

If your conjecture is different from the computer-generated rotation, explain why.



Description of transformation:
Rotate the flag 135° about the
marked center.

If your conjecture is different from the computer-generated rotation, explain why.



Flag used with permission from Patrick Thompson

You Decide

Ms. Smith hired Mr. Jones to work at the Lose-It-All Casino. Ms. Smith made the following salary offer to Mr. Jones. He can receive \$400 per week, or he can determine his weekly salary by the following game of chance:

> Draw three bills out of a bag containing six bills. The bag contains a \$5, a \$10, a \$20, a \$50, a \$100, and a \$500 bill. The value of the three bills will be Mr. Jones weekly salary for the year.

Which offer should Mr. Jones accept?

Before answering this question, let's look at a simpler question. Suppose Mr. Jones were to draw two bills out of a bag containing a \$20, a \$50, a \$100, and a \$500 bill. We could list all the possible outcomes.

\$500	\$500	\$500
\$100	\$ 50	\$ 20

\$100	\$100
\$ 50	\$ 20

\$ 50
\$ 20

Half the outcomes are grater than \$400, and half the outcomes are below \$400. Does this help Mr. Jones make a decision?

Let's analyze this further. The *expected value* is the average amount of money you would draw from the bag each time if you played the game many, many times. We find the expected value by multiplying the value of each outcome times the probability of that outcome; then adding all these products together. Try it for the four-bill game.

The probability for each draw is $^{1}/_{6}$ so the expected value would be

$^{1}/_{6}\$500 + \$100) + {}^{1}/_{6}(\$500 + \$50) + {}^{1}/_{6}(\$500 + \$20) +$
$^{1}/_{6}\ (\$100 + \$50) + {}^{1}/_{6}(\$100 + \$20) + {}^{1}/_{6}(\$50 + \$20) = \$335$

We can now return to the six-bill game. First, list all the possible outcomes when drawing three bills from the bag.

\$500	\$500	\$500	\$500
\$100	\$100	\$100	\$100
\$ 50	\$ 20	\$ 10	\$ 5

\$500	\$500	\$500
\$ 50	\$ 50	\$ 50
\$ 20	\$ 10	\$ 5

\$500	\$500
\$ 20	\$ 20
\$ 10	\$ 5

\$500
\$ 10
\$ 5

\$100	\$100	\$100
\$ 50	\$ 50	\$ 50

186

$ 20

$ 10

$ 5

$100
$ 20
$ 10

$100
$ 20
$ 5

$100
$ 10
$ 5

$ 50
$ 20
$ 10

$ 50
$ 20
$ 5

$ 50
$ 10
$ 5

$ 20
$ 10
$ 5

How many different outcomes are possible?

What is the probability of each outcome?

What is the probability of drawing more than $400?

What is the expected value?

Based on the probability and expected value, should Mr. Jones take the $400 salary or should he play the game and let the draw determine his salary?

Are there any other factors that Mr. Jones should consider before making this decision?